ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA

DIPARTIMENTO DELL'ENERGIA ELETTRICA E DELL'INFORMAZIONE
"Guglielmo Marconi"
DEI
Corso di Laurea Magistrale in Ingegneria Elettronica

Tesi di laurea
in
Sistemi Elettronici ad Alta Affidabilità

SISTEMI DI REGOLAZIONE DELL'ALIMENTAZIONE AFFIDABILI PER PROCESSORI MULTI-CORE

Candidato:
Enrico Vicini

Relatore:
Chiar.ma Prof.ssa Cecilia Metra

Correlatori:
Prof. Martin Omaña
Dott. Ing. Stefano Petrucci
Dott. Ing. Giuseppe Froio
Dott. Ing. Mauro Pipponzi

Anno Accademico 2018/2019
Sessione II

Ringraziamenti

Desidero partire dalla fine e riavvolgere il nastro di questo percorso per ringraziare tutte quelle persone che hanno contribuito non solo alla mia formazione professionale, ma anche alla mia crescita personale durante la mia esperienza accademica.

Innanzitutto, il primo immenso e doveroso ringraziamento va alla Prof.ssa Cecilia Metra, che mi ha dato la possibilità di mettermi alla prova su un tanto complesso quanto interessante progetto di tesi, in collaborazione con un'azienda tra le grandi protagoniste dell'evoluzione tecnologica cui continuiamo ad assistere. La ringrazio, inoltre, per avermi sempre coinvolto nei vari seminari da Lei organizzati esponendomi a tematiche interessanti, nonché di avermi permesso di collaborare all'organizzazione dei due eventi IEEE in dicembre. In secondo luogo, un altrettanto grande ringraziamento va al Prof. Martin Omaña per la continua disponibilità ed il costante supporto nei miei confronti. Gli sono grato per avere sempre ascoltato le mie idee, indipendentemente dal fatto che potessero essere valide o meno, dimostrandosi aperto ad un confronto costruttivo, da cui raccogliere preziosi suggerimenti.

Passo ora a ringraziare i miei compagni di avventura a Bologna, nonché coinquilini, Luca e Thomas, che hanno saputo rendere la vita divertente in quell'appartamento gelido (o forse era solo la mia camera) ed eccessivamente vetusto nel quale passavamo la settimana. Li ringrazio per i viaggi in macchina, le pulizie, l'amatriciana, il tiramisù e l'Oki di Thomas, ed i lampi di genio delle undici di sera di Luca, ... e per aver sopportato il mio carattere, talvolta difficile.

Voglio ringraziare gli "zii di Bologna" per le prime due settimane a Bologna, che ho trascorso a casa loro sentendomi come a casa mia. Durante quel periodo ho avuto modo di conoscerli meglio, molto più di quanto li conoscessi fino ad allora per le poche occasioni che avevamo di vederci data la distanza. Li ringrazio non solo per avermi ospitato nel primo periodo, ma anche per avermi introdotto a questa (per me) nuova città, nonché per le sempre ottime cene.

Un ringraziamento speciale va a Giacomo, sicuramente la persona che più mi è mancata dalla mia precedente esperienza triennale a Trento. In quei tre anni abbiamo avuto modo di instaurare un legame che va oltre la parentela. Lo ringrazio per le lunghe chiamate e chiacchierate, in cui è sempre disponibile ad ascoltare e soprattutto capire.

Nel mio cambio di percorso ho dovuto lasciare a Trento un amico con cui ho condiviso tanti momenti durante la triennale: lezioni, partite a carte nelle pause, qualche corsa al Gocciadoro, nonché le cene a casa mia quando Giacomo non c'era ed io non avevo voglia di farmi da mangiare. Nonostante la distanza, siamo comunque riusciti a fare in modo di vederci. Lo ringrazio per la solida amicizia che abbiamo costruito e per il cibo altoatesino che porta quando mi viene a trovare.

L'ultima persona che ringrazio di Trento è Roberto, che da tutor della tesi triennale è diventato per me una sorta di mentore.

Un ringraziamento particolare per una persona veramente tanto speciale: Laura. Le sono grato per essermi stata sempre vicina durante le scelte, il cui peso diventa sempre maggiore, e per sapermi ascoltare. La ringrazio per le sue tante idee e per la sua spensieratezza, che senza dubbio sortiscono continuamente un effetto positivo sulla mia persona. Spero che possa continuare ad essere al mio fianco.

Ringrazio le vere amicizie che ho conservato dal liceo: Sebastiano, Matteo e Serena, pochi, ma buoni, come si è soliti dire. Nonostante ciascuno di noi abbia intrapreso strade diverse e le occasioni per vedersi siano sempre più rare (ad eccezione di Sebastiano per ovvie ragioni), ogni volta che ci ritroviamo sembra che il tempo non sia passato.

Prima di terminare, probabilmente devo ringraziare il primo ingegnere che ho conosciuto: lo zio Federico. Ancora oggi non ho ben capito cosa abbia fatto nascere in me la passione e l'interesse per l'elettronica e l'informatica, ma sicuramente parte del merito spetta a lui.

Infine, voglio ringraziare i miei genitori, che mi hanno sempre supportato nel mio percorso. Li ringrazio per avermi incoraggiato nei momenti di difficoltà e per avere la pazienza di ascoltare le mie preoccupazioni, sapendomi sempre consigliare nel modo giusto nelle mie scelte. Devo a loro in primis la persona che sono oggi, sia negli aspetti positivi sia in quelli negativi (perché è giusto che si assumano la responsabilità di entrambi).

Indice

Introduzione

Motivazioni

Con il progredire della tecnologia microelettronica, sempre più il limite alle prestazioni computazionali dei moderni SoCs è costituito dal consumo di potenza, in modo particolare per quei sistemi alimentati a batteria [1–3]. Per aumentare l'efficienza energetica e ridurre i consumi sono state messe a punto diverse strategie di power management, ad esempio la Dynamic Voltage and Frequency Scaling (DVFS) [3], con lo scopo di aumentare la performance-per-watt nei SoCs complessi. Queste tecniche fanno tipicamente uso di una Power Control Unit (PCU) che monitora in real-time l'attività dei blocchi che compongono il SoC (ad esempio, i cores, le caches, la parte di I/O). L'introduzione della PCU ha permesso di controllare i power gates di ciascun power domain, in modo da ridurre drasticamente il consumo di potenza dovuto al leakage in quei domini in stato idle [1]. Tuttavia, ogni dominio attivo era costretto ad operare con la tensione più alta fra tutte quelle richieste. Fornire a ciascun dominio la tensione ottimale, sulla base del monitoraggio effettuato dalla PCU, avrebbe richiesto l'implementazione di più Voltage Regulators (VRs) sulla motherboard. Questa soluzione risultava inappropriata a causa dell'aumento di costi ed area, oltre alla necessità di avere pin extra sul package.

La recente integrazione dei VRs direttamente nel die ha dato la possibilità di fornire la tensione ottimale a ciascun dominio sulla base della propria attività, monitorata dalla PCU. In questo caso si parla di Fully Integrated Voltage Regulators (FIVRs). Infatti, i FIVRs non richiedono pin di I/O supplementari sul die (al contrario di VRs implementati sulla motherboard), possono essere posizionati in prossimità dei corrispettivi carichi, in modo da minimizzare la potenza dissipata a causa delle interconnessioni, e consentono transizioni più rapide nel variare la tensione fornita al dominio (denominata power state) [1–3]. Questo non solo ha permesso di migliorare l'efficienza energetica in termini di riduzione della tensione per quei domini con un carico di lavoro moderato, ma ha anche consentito di aumentare la potenza disponibile per un temporaneo

aumento delle prestazioni. Si consideri, ad esempio, il tipico caso di un processore multi-core in cui solo un core è soggetto ad un elevato carico computazionale: la diminuzione della potenza assorbita dagli altri core può essere sfruttata per fornire un boost prestazionale al core attivo. L'introduzione del FIVR ha permesso di implementare una tecnica di power management nota come *per-core p-state* (PCPS), che ha sostituito la precedente *per-socket p-state* (PSPS), la quale forzava tutti i cores all'interno del medesimo socket allo stesso power state, determinato dal core richiedente il p-state più alto [2].

Nei moderni SoCs, i FIVRs sono implementati come convertitori switching DC/DC di tipo Buck [1–3]. Il principio di funzionamento del convertitore Buck prevede che vi siano due interruttori (di cui almeno uno controllato) che si chiudono ed aprono per generare una forma d'onda rettangolare con un valore medio pari a quello desiderato. Un filtro passa-basso LC elimina tutte le armoniche del segnale rettangolare ad eccezione della continua. Nel FIVR è presente, inoltre, il FIVR Control Module (FCM), che comprende un circuito di Pulse Width Modulation (PWM), il quale confronta la tensione di uscita del FIVR con il riferimento desiderato, un convertitore digitale-analogico (DAC), che converte la parola di riferimento fornita dalla PCU in una tensione "analogica", ed un blocco di compensazione, che permette di ridurre il tempo di risposta del FCM in presenza di variazioni nell'assorbimento di corrente da parte del carico.

Un problema rilevante che affligge i FIVRs è legato alla loro affidabilità. Infatti, poiché la maggior parte dei componenti che costituiscono il FIVR sono realizzati sul die (ad eccezione degli induttori e dei condensatori sul package), essi sono affetti da guasti e fenomeni di aging, la cui entita è sempre più accentuata al progredire dello scaling tecnologico. Questi problemi possono compromettere il corretto funzionamento del FIVR.

L'uscita di anche solo uno dei parametri significativi del FIVR da un range di tolleranza determinato a priori risulta assolutamente non accettabile in contesti critici. Un esempio importante è costituito dal settore dell'automotive, che negli ultimi anni, in virtù del crescente interesse, sta conoscendo un rapido sviluppo. In questo campo, in cui sempre più le tecnologie di assistenza alla guida, guidate dall'intelligenza artificiale, assumono il controllo del veicolo (*autonomous drive*), vi è esigenza di elevata *reliability*, cioè avere la garanzia che il sistema continui a funzionare correttamente. Infatti, un eventuale *fault* del sistema risulterebbe catastrofico. Si consideri, ad esempio, un guasto del FIVR tale da abbassare la tensione rispetto a quella richiesta da un core che sta svolgendo operazioni atte a prendere una decisione sul compor-

tamento del veicolo in una specifica situazione. L'abbassamento della tensione causerebbe un rallentamento del core tale da far sì che vengano generati errori logici portando al malfunzionamento del sistema con un conseguente potenziale rischio per le persone all'interno dell'abitacolo.

Per verificare le prestazioni dei FIVRs, alcuni SoCs implementano tecniche di *design for testability* (DFT). Tuttavia, la caratterizzazione dei parametri chiave del FIVR non può essere effettuata per ogni chip fabbricato, a causa dell'elevato costo in termini di tempo impiegato, bensì viene performata solo su alcuni esemplari. In ogni caso, anche se tale verifica venisse effettuata su ogni chip, questa non garantirebbe il corretto funzionamento del FIVR sul campo (e delle tecniche di power management annesse), per il quale sono, invece, richieste delle specifiche tecniche di on-line testing.

Per ottenere la *safety* del sistema, è necessario che il monitor sia in grado di rivelare anche i guasti al proprio interno o che sia in grado di continuare a funzionare correttamente nonostante il verificarsi di un guasto al proprio interno. Nell'ambito dell'*autonomous drive*, il documento che illustra tutti i requisiti di *safety* è lo standard ISO 26262, che costituisce il riferimento nel settore.

Lavoro svolto

Questa tesi si svolge nell'ambito di un progetto di ricerca in collaborazione con *Intel Corporation*.

Il lavoro è partito dai risultati ottenuti in [4, 5] e propone un nuovo schema a basso costo per la rivelazione di guasti che possono verificarsi durante il normale funzionamento sul campo del FIVR. Il monitor si differenzia da quello in [5] e da quello suggerito nello standard ISO 26262 per l'utilizzo di un approccio completamente diverso, basato su un confronto di ritardi di propagazione, anziché su un confronto direttamente effettuato tra tensioni. Lo schema di rivelazione è stato progettato in modo da essere self-checking, cioè è in grado di auto-collaudarsi per rivelare guasti al proprio interno o continuare a funzionare correttamente nonostante la presenza di guasti interni.

La tesi è suddivisa in cinque capitoli, così suddivisi:

- il capitolo 1 presenta le nozioni di base inerenti il FIVR, i guasti più probabili che possono verificarsi al suo interno ed i fenomeni di aging, mostrando inoltre lo schema di rivelazione dei guasti proposto dallo standard ISO 26262

- il capitolo 2 propone una nuova strategia di rivelazione dei guasti, discutendone il funzionamento ad alto livello

- il capitolo 3 illustra una possibile implementazione del monitor a livello elettrico
- il capitolo 4 mostra i risultati di alcune simulazioni per dimostrare il funzionamento del monitor
- il capitolo 5 riporta i costi del monitor in termini di area occupata, nonché l'analisi della self-checking abiliy dello schema di rivelazione

Capitolo 1

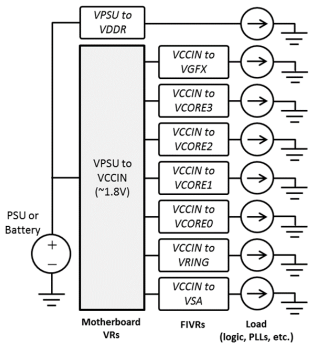
Concetti preliminari

Obiettivo di questo capitolo è presentare il funzionamento e riportare le prestazioni del FIVR in [1]. In seguito vengono illustrati i modelli di guasto più probabili che possono affliggere il FIVR, che includono quelli di transistor stuck-open, transistor stuck-on e bridging resistivo, quindi vengono riportati sinteticamente i risultati dell'analisi dei guasti condotta sul FIVR contenuta in [5]. Successivamente viene trattato il fenomeno dell'aging, mostrando quanto evidenziato in [5] circa l'effetto dell'invecchiamento sui transitori del FIVR. Infine, viene illustrata la soluzione proposta dallo standard ISO 26262 per rivelare i guasti che possono affliggere il FIVR.

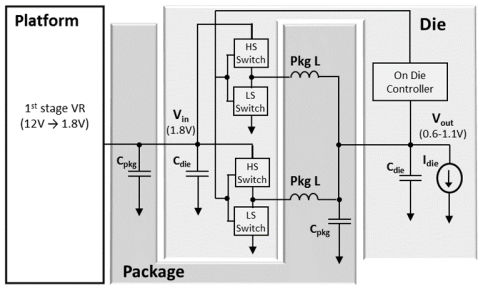
1.1 Fully Integrated Voltage Regulator

La quarta generazione dei processori Intel Core, denominata Haswell ed introdotta nel 2014, è alimentata da FIVRs [1]. Questi dispositivi sono integrati direttamente nel die del microprocessore, realizzato in tecnologia a $22nm$, e sono implementati come regolatori Buck multifase (fino sedici fasi) a $140MHz$, aventi frequenza a guadagno unitario pari a $80MHz$. I FIVRs sono altamente configurabili e sono impiegati per potenze dai $3W$ dei tablets fino ai $300W$ dei servers. L'introduzione del FIVR ha permesso di incrementare la durata della batteria anche più del 50% e di aumentare anche più del doppio la potenza di picco disponibile per carichi di lavoro onerosi.

In figura 1.1a viene riportato lo schema a blocchi dei domini alimentati dai FIVRs. Il primo stadio è costituito da un VR sulla scheda madre che abbassa la tensione fornita dall'alimentatore o dalla batteria (in un range da 12 a $20V$) a circa $1.8V$, che viene distribuita nel die. Il secondo stadio di conversione è composto di un certo numero di FIVRs (da otto a trentuno a seconda del prodotto), utilizzati per alimentare i vari domini.



(a) Schema dei domini alimentati dai FIVRs.



(b) Schema di un FIVR a due fasi.

Figura 1.1: [1]

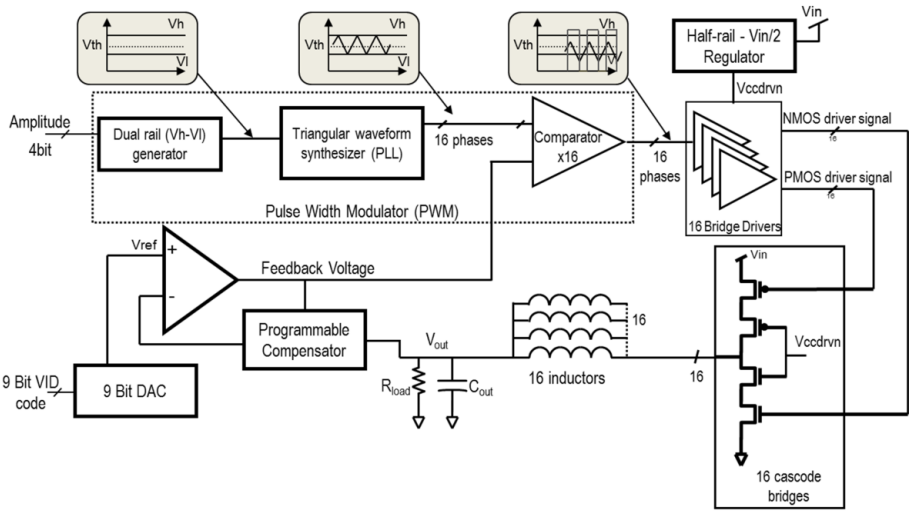


Figura 1.2: Schema a blocchi del singolo FIVR. [1]

Ciascun FIVR viene programmato indipendentemente dalla PCU, che stabilisce una serie di parametri per minimizzare il consumo di potenza totale del die.

Una rappresentazione schematica di un singolo FIVR (a due fasi) viene riportato in figura 1.1b. Come si può notare, i transistor di potenza, il circuito di controllo e i condensatori di disaccoppiamento ad alta frequenza sono sul die, mentre gli induttori ed i condensatori di disaccoppiamento a media frequenza di ingresso stanno sul package.

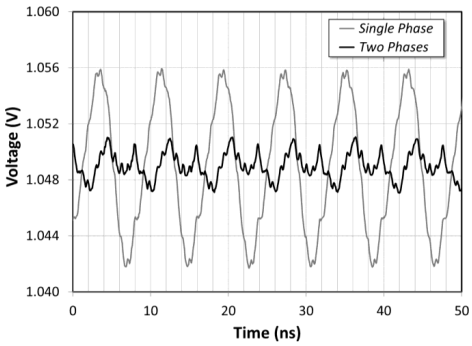
Il funzionamento del FIVR è illustrato più nel dettaglio in figura 1.2. I transistori di potenza del convertitore Buck sono sostituiti dalla configurazione a cascode di transistori nMOS e pMOS. La configurazione a cascode permette di implementare gli switches di potenza con nor-

mali transistori a $22nm$ pur avendo una tensione di alimentazione di $1.8V$. Due trasistori sono pilotati da bridge drivers che supportano zero-voltage-switching (ZVS) e zero-current-switching (ZCS), mentre gli altri due sono connessi a V_{ccdrvn}, regolata a $V_{in}/2$. Tale tensione è anche la alimentazione positiva del nMOS bridge driver e l'alimentazione negativa del pMOS bridge driver. I segnali `phases` sono generati da un PWM. Il segnale di feedback fornito al comparatore del PWM viene generato comparando la tensione di uscita del FIVR (attraverso una rete compensatrice) con la tensione di riferimento, ottenuta dalla conversione in analogico di una parola di nove bits fornita dalla PCU.

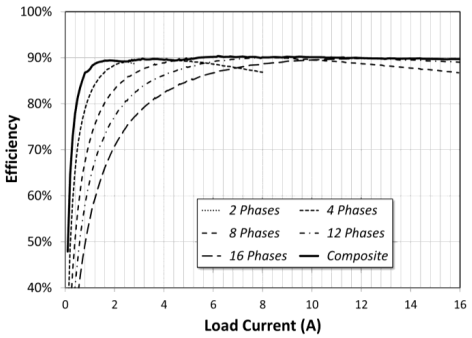
Ciascun FIVR è controllato da un proprio FCM, che include il circuito che genera i segnali di PWM ed il compensatore programmabile. La frequenza del PWM, il guadagno, l'attivazione della fase e l'angolo di ogni fase sono programmabili al fine di raggiungere un'elevata efficienza ed un ripple minimo. Il compensatore programmabile fornisce un'elevata larghezza di banda, che si traduce in un'elevata velocità di risposta alle variazioni di assorbimento di corrente da parte del carico. Il compensatore di ogni dominio è programmabile individualmente in funzione del proprio filtro di uscita e può essere riprogrammato mentre il dominio è attivo, al fine di mantenere una risposta ottimale.

Per minimizzare le perdite del FIVR, la PCU configura dinamicamente ciascun FCM in base all'attività, specificando la tensione di uscita per raggiungere la frequenza desiderata. Inoltre, determina il numero di fasi attive e programma la rete compensatrice. Queste operazioni permettono ad ogni FIVR di operare con un'efficienza prossima a quella massima in un vasto range di condizioni di carico.

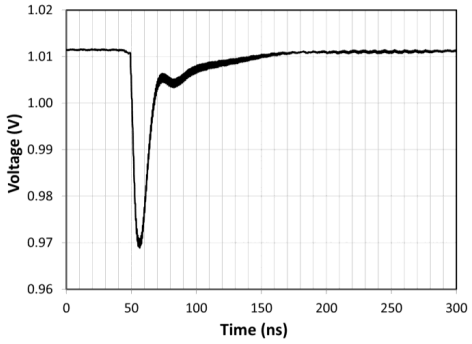
Per quanto concerne le prestazioni del convertitore dichiarate, vengono di seguito riportate alcune significative figure di merito. In figura 1.3a viene confrontato il ripple di tensione di un dominio alimentato a due fasi rispetto a una sola: si può osservare come, nel caso di due fasi, il ripple sia inferiore a $4mV$, quindi inferiore all'1% del set point. In figura 1.3b viene, invece, mostrata l'efficienza misurata per un singolo dominio al variare della corrente richiesta dal carico: si può notare come si riesce sempre a mantenere un'efficienza intorno al 90% variando il numero di fasi attive. In particolare, all'aumentare della corrente richiesta dal carico, si rende necessaria l'attivazione di un numero crescente di fasi. Un'altra figura di merito rilevante è riportata in figura 1.3c. Il grafico rappresenta l'andamento della tensione nel tempo al verificarsi di uno step di corrente richiesta dal carico pari a $8.5A$: la combinazione di un anello di retroazione a banda larga e di condensatori di disaccoppiamento direttamente sul die consentono di limitare la caduta di tensione a meno di $50mV$, nonostante lo step di corrente inferiore a $1ns$, e di



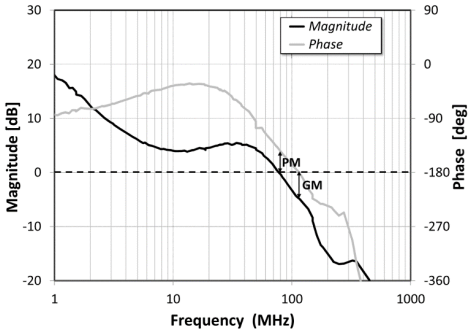
(a) Ripple al variare del numero di fasi.



(b) Efficienza al variare della corrente di carico.



(c) Risposta in presenza di uno step di corrente.



(d) Diagramma di Bode.

Figura 1.3: Figure di merito del FIVR. [1]

riportare la tensione al valore nominale in meno di $100ns$. Infine, in figura 1.3d viene mostrata la funzione di trasferimento ad anello aperto del FIVR, evidenziando, in particolare, la frequenza a guadagno unitario pari a $78MHz$, in corrispondenza della quale si ha un margine di fase di $40°$.

In tabella 1.4 si riporta il confronto con alcuni lavori precedenti. In particolare, si osserva che il FIVR opera ad una frequenza di commutazione più alta, in parte anche in virtù della bontà della caratteristica della carica di gate dei transistori, che permette di ottenere un'efficienza fino al 90%. Questo si traduce in un aumento della durata della batteria nei dispositivi mobili superiore al 50% e in un incremento del picco di potenza disponibile di oltre il doppio. Inoltre, non dovendo utilizzare molteplici VRs sulla scheda madre, è possibile ridurre le dimensioni della stessa oppure sfruttare lo spazio disponibile per introdurre nuove funzionalità.

Parameter	G. Schrom et al., 2010 [2]	T. DiBene et al., 2010 [3]	N. Sturcken et al., 2012 [4]	This Work
Process node	130 nm	90 nm	45 nm	22 nm
Switching Frequency	60 MHz	100 MHz	80 MHz	140 MHz
Unity Gain Freq	5 MHz	Not Published	Not Published	80MHz
Efficiency	85-88%, 3.3V:1.0V	76%	83%, 1.5V:1.0V	90%, 1.7V:1.05V
Total Output Imax capability	50 A	Limited by first stage and thermals (Up to 400 A)	1.2 A	Limited by first stage and thermals (Up to 700 A)
Imax/VR die area	1.3 A/mm^2	8 A/mm^2	1.7 A/mm^2	31 A/mm^2
Voltage rail count	4	20	1	8 to 31
Phase count	16	320	4	49 to 360
Integration level	MCM[a]	MCM[a]	Integrated into network die	Integrated into CPU die
Inductor technology	Package trace, & magnetic discrete	Magnetic thin-film on VR die	Discrete wire-wound air core	2D array of package trace
Capacitor type	Ceramic package caps	Ceramic package caps	Die Cap	Die Cap - MIM
Cout per Max Amp	2000 nF/A	not published	15 nF/A	7 nF/A

[a] MCM – Multi Chip Module – the active circuitry is on a separate die assembled on the same package

Figura 1.4: Confronto con lavori precedenti. [1]

1.2 Guasti probabili nel FIVR

Come ogni dispositivo che viene fabbricato, anche il FIVR è soggetto ad una fase di collaudo. Poiché il numero di difetti fisici (ad esempio piste rotte o cortocircuitate) che potrebbero occorrere su un chip è molto elevato, il collaudo è basato su alcuni *modelli di guasto* che descrivono l'effetto che più difetti fisici hanno sul comportamento (statico e/o dinamico) del circuito, riducendo così la molteplicità di guasti da considerare.

Il processo di collaudo deve essere il più rapido possibile, dato che contribuisce al time-to-market, e deve essere il più accurato possibile, onde evitare di scartare chip correttamente funzionanti (perdita di resa) o di immettere sul mercato chip malfunzionanti.

Vengono di seguito rapidamente illustrati i modelli di guasto più probabili per quanto riguarda il FIVR oggetto di questa tesi.

1.2.1 Bridging resistivo

I guasti di tipo bridging resistivo (BF) descrivono quei difetti fisici il cui effetto sul comportamento logico del circuito è far sì che il circuito si comporti come se due nodi o due linee fossero collegati da un cammino resistivo di resistenza pari a quella di bridging. Il valore di resistenza di bridging è diverso a seconda del difetto fisico, per avere adeguata aderenza del modello con la realtà fisica.

La presenza di un guasto di questo tipo fa sì che per particolari configurazioni degli ingressi possa esserci un cammino conduttivo dall'alimentazione alla massa. Questo porta, innanzitutto, ad avere l'uscita indeterminata (il cui livello dipende dall'entità della resistenza di bridging rispetto alla conducibilità dei transistor), oltre ad avere un impatto sulle prestazioni dato sia dalla dissipazione di potenza statica sia da

un rallentamento dei transitori per il gate logico a valle, i cui transistori pilotati dal circuito in esame si ritrovano una minore tensione di overdrive.

Se un guasto di tipo BF determini un errore logico dipende dalla soglia logica del gate a valle. Si ha errore logico se il valore di resistenza di bridging è sufficientemente piccolo. Tale valore può essere individuato effettuando una simulazione DC parametrica al variare della resistenza e visualizzando l'uscita del gate logico a valle.

È possibile rivelare il guasto mediante confronto tra il valore logico prodotto dall'uscita ed il valore atteso, se il valore di tensione intermedio prodotto sta, rispetto alla soglia logica del gate a valle, dalla parte opposta rispetto al valore atteso. Tramite I_{DDQ} testing si può misurare la corrente statica che attraversa il circuito, tuttavia questo limita la frequenza massima a cui si può effettuare il collaudo (si deve assicurare che i transitori siano esauriti) e risulta inefficace con lo scaling della tecnologia. L'inefficacia è dettata dal fatto che, con l'avanzare dello scaling tecnologico, da un lato la diminuzione della tensione di soglia fa sì che aumenti la corrente di leakage, dall'altro il numero di dispositivi per unità di area aumenta, per cui aumenta complessivamente la corrente statica che attraversa il circuito, rendendo difficile l'individuazione di una soglia che consenta di distinguere tra chip guasti e correttamente funzionanti. Può essere altrimenti rivelato durante il testing at-speed, date le minori prestazioni dinamiche del gate a valle.

Per determinare il range di valori assunti dalla resistenza di bridging da considerare nell'analisi del FIVR, in [5] è stato implementato un invertitore simmetrico a $22nm$ con l'uscita collegata all'alimentazione V_{DD} tramite una resistenza R_B. Ponendo in ingresso un valore logico alto (vettore di attivazione del guasto), si è osservato che per $R_B \in [0, 18k]\Omega$ si ottiene in uscita un valore logico errato, mentre per $R_B \in (18k, 300k]\Omega$ in uscita si ha il valore logico corretto. Questi ultimi risultano essere i guasti più pericolosi, perché non rivelabili tramite semplici tecniche di collaudo nei confronti di guasti di tipo stuck-at (SA). Si è ritenuto, quindi, sufficiente considerare valori $R_B \in [0, 100k]\Omega$, poiché per valori di resistenza superiori, la tensione di uscita decresce molto lentamente tendendo ad un valore costante.

1.2.2 Transistor Stuck-On

I guasti di tipo stuck-on (SON) descrivono quei difetti fisici il cui effetto sul comportamento del circuito è di far sì che il circuito si comporti come se un transistor fosse sempre in conduzione (indipendentemente dal valore della tensione applicata al gate). Può essere visto come un guasto di tipo BF di resistenza pari alla resistenza equivalente che esibisce il

transistor in conduzione, quindi le considerazioni sono del tutto analoghe a quelle fatte alla sezione 1.2.1.

1.2.3 Transistor Stuck-Open

I guasti di tipo stuck-open (SOP) descrivono quei difetti fisici il cui effetto sul comportamento del circuito è di far sì che il circuito si comporti come se un transistor fosse sempre spento (indipendentemente dal valore della tensione applicata al gate).

Se viene applicata in ingresso una configurazione in grado di attivare il guasto, l'uscita si ritrova in alta impedenza, in quanto non collegata né all'alimentazione né alla massa. In questo caso rimane in uscita il valore precedentemente caricato ed il circuito si comporta come sequenziale anziché combinatorio. Per rivelare il guasto sono allora necessari due vettori di test: il primo di *inizializzazione* serve per portare l'uscita al valore opposto rispetto a quello dato dalla rete in cui è presente il transistor guasto, il secondo di *attivazione* per cercare di commutare l'uscita attraverso un percorso che includa il transistor guasto. Il guasto viene rivelato attraverso la non commutazione dell'uscita.

1.3 Effetti dei guasti più probabili sul FIVR

Si riporta di seguito una panoramica degli effetti dei guasti più critici e delle loro probabilità, come presentate in [5]. La criticità dei blocchi viene valutata in funzione dei seguenti indicatori:

- P^{blocco}: probabilità del failure mode in analisi rispetto al numero di guasti considerati nel blocco in esame
- P^{FIVR}: probabilità del failure mode in analisi rispetto al numero di guasti considerati in tutto il FIVR

Come si evince dalla tabella 1.1, tra tutti i blocchi che costituiscono il FIVR, il più critico risulta essere il DAC, poiché i suoi guasti hanno un'alta probabilità di verificarsi (P^{FIVR}).

Per riassumere:

- il 40.9% dei guasti considerati non causa alcun effetto sull'uscita del FIVR
- il 55% dei guasti considerati fa sì che l'uscita si discosti dal valore di riferimento per più del 5%
- il 3.4% dei guasti causa un effetto catastrofico sull'uscita del FIVR, tuttavia questi guasti sono facilmente rivelabili dalla PCU, poiché il core alimentato dal FIVR guasto non funziona

Blocco	Failure mode	P^{blocco} [%]	P^{FIVR} [%]	Effetto su V_{out}
Single Rail Generator	Errore di linearità con curva di conversione monotona	60.5	30.6	Aumento del ripple
	Curva di conversione non monotona	44	11	Aumento del ripple
	Errore di guadagno	21	5.3	No effetti
	O_{RAIL} in stuck-at	11	2.7	Valore costante errato
	Errore di offset	25	6.2	Aumento del ripple
Triangular Waveform Synthesizer	Frequenza errata	43	5.5	Aumento del ripple
	Swing errato	25	1.6	Aumento del ripple
	O_{TRI} in stuck-at	25	1.6	Valore costante errato
Comparatore	Range errato	44	0.9	Valore costante errato; Range di variazione errato
	O_{TRI} in stuck-at	33	0.7	Valore costante errato
9 bits DAC	Errore di linearità con curva di conversione monotona	74	71.3	Range di variazione errato
	Curva di conversione non monotona	72	34.7	Range di variazione errato
	Errore di guadagno	16	8.5	Range di variazione errato
	V_{ref} in stuck-at	2	0.9	Valore costante errato; Aumento del ripple
	Errore di offset	18	8.7	Range di variazione errato
Compensatore programmabile	Attenuazione errata	65.38	3.9	Valore costante errato; Range di variazione errato; Aumento del ripple
Bridge driver	Segnale costante	42	5.2	Range di variazione errato;
	Variazione ampiezza	51	6.2	Aumento del ripple

Tabella 1.1: Possibili guasti del FIVR. [5]

1.4 Modelli di aging

La degradazione delle caratteristiche del transistore MOS a causa dell'aging è diventato critico per le tecnologie scalate a partire dai $45nm$ [6, 7]. L'aging degrada le prestazioni del circuito nel tempo, accorciando il tempo di vita del circuito stesso, ed introduce potenziali guasti sul campo.

L'aging è dovuto principalmente alla degradazione nel tempo del dielettrico di gate e dell'interfaccia tra il dielettrico ed il Silicio. Due importanti meccanismi che concorrono in tale degradazione prendono il nome di *Hot Carrier Injection* (HCI) e *Bias Temperature Instability* (BTI) [6–8]. Questi meccanismi sono particolarmente rilevanti nei nodi tecnologici in cui il dielettrico di gate ha uno spessore equivalente di poche molecole e con l'utilizzo di transistori high-K metal-gate. Per quantificare la degradazione delle performances del circuito nel tempo è richiesta una fase di testing estremamente lunga e costosa, che aumenta complessivamente i costi di produzione. In alternativa, i designers sovradimensionano i circuiti critici, ma questo determina un aumento del costo del chip. Si rende, dunque, essenziale un metodo conveniente per stimare il tempo di vita di un circuito, specialmente per le applicazioni caratterizzate da un'esigenza di elevata reliability, come ad esempio il settore dell'automotive.

Il simulatore commerciale HSPICE include un tool denominato *MOS Reliability Analysis* (MOSRA) che permette di predire l'effetto dell'aging sulle prestazioni del circuito con i modelli forniti o con modelli caricati dall'utente.

L'aging del dispositivo è il risultato di una continua degradazione delle sue caratteristiche, a causa dello stress elettrico cui è sottoposto. Il tool MOSRA sfrutta un modello di aging per tradurre un certo ammontare di stress elettrico in una determinata degradazione del dispositivo. Tipicamente questi modelli dipendono dalle condizioni operative del dispositivo (cioè tensioni, correnti e temperature) nonché dalle sue dimensioni. La degradazione risultante può essere applicata in due diversi modi:

- l'ammontare di stress viene convertito nella degradazione di parametri chiave del MOSFET (cioè tensione di soglia, mobilità, ...)

- lo stress viene direttamente convertito in una degradazione delle caratteristiche del dispositivo (cioè una degradazione percentuale della corrente di drain e della sua conduttanza)

Il secondo approccio possiede l'innegabile vantaggio di essere semplice, tuttavia il primo permette di separare i differenti effetti che contribuiscono alla degradazione del dispositivo. Ciò comporta una migliore accuratezza nel calcolo della degradazione della corrente e della conduttanza su un vasto range di tensioni di polarizzazione.

1.4.1 Hot Carrier Injection

In presenza di campi elettrici intensi, parte della carica mobile nel canale dalla parte del drain (elettroni per i transistori nMOS, lacune per i pMOS) viene intrappolata nel dielettrico di gate, cambiandone le proprietà elettriche nel tempo [8]. Il fatto che un portatore, anziché contribuire alla conduzione muovendosi normalmente nel canale o ricombinarsi nel substrato, si sposti nel dielettrico per effetto tunnel, determina un aumento della corrente di leakage di gate ed un possibile danneggiamento alla struttura atomica del dielettrico. Il fenomeno di HCI provoca un innalzamento della tensione di soglia, quindi la commutazione del transistor viene rallentata.

L'aggettivo "hot" non si riferisce alla temperatura fisica del substrato, bensì al termine di temperatura efficace utilizzato nel modellare la densità dei portatori secondo la distribuzione di Fermi-Dirac. Il termine era stato inizialmente introdotto per descrivere i portatori non in equilibrio nei semiconduttori.

Il modello HCI di MOSRA tiene in considerazione le dipendenze dalla polarizzazione del dispositivo, nonché la temperatura cui è operativo, risultando accurato in un vasto range di lunghezze di canale e per diversi spessori del dielettrico di gate.

1.4.2 Bias Temperature Instability

Il fenomeno di BTI si distingue in *negative* BTI (NBTI) per i transistori pMOS e *positive* BTI (PBTI) per i transistori nMOS. In entrambi i casi, non è necessario che scorra della corrente tra source e drain [8].

Di seguito viene trattato il fenomeno del NBTI, ma considerazioni analoghe valgono per il PBTI. Il NBTI è il risultato della combinazione di cariche positive intrappolate nel dielettrico e cariche intrappolate all'interfaccia tra dielettrico e Silicio in presenza di una tensione di gate-source negativa, soprattutto in caso di temperatura elevata [9]. L'ammontare delle cariche positive intrappolate nel dielettrico è circa uguale alla quantità di cariche intrappolate all'interfaccia. L'effetto del NBTI è quello di aumentare la tensione di soglia, quindi di ridurre la conducibilità del transistor. L'entità del fenomeno di NBTI è sicuramente maggiore per i transistori pMOS per due motivi principali: innanzitutto, nelle tipiche condizioni di funzionamento di un circuito CMOS, solo

i transistori pMOS sono sottoposti a tensioni di gate-source negative; inoltre, nel caso di transistori pMOS, le cariche attirate all'interfaccia sono positive, quindi l'effetto si somma alle cariche positive intrappolate nel dielettrico.

Il fenomeno di NBTI richiede sempre maggiore considerazione nel progettare un circuito per diversi motivi. Innanzitutto, le tensioni di alimentazione non assistono ad uno scaling così veloce come quello cui è soggetto lo spessore del dielettrico di gate, risultando in campi elettrici più intensi, quindi un'accentuazione del fenomeno. In secondo luogo, in presenza di minori tensioni di alimentazione, a parità di aumento della tensione di soglia a causa dell'NBTI, la degradazione della conducibilità del transistor è maggiore. Infine, l'introduzione di Azoto nel dielettrico di gate con lo scopo di ridurre la corrente di leakage che attraversa il gate per effetto tunnel ha l'effetto collaterale di aumentare il fenomeno di NBTI. Inoltre, con l'aumentare dello scaling tecnologico, aumenta la variabilità del fenomeno da transistor a transistor, in quanto l'effetto risulta mediato su un'area di gate progressivamente inferiore.

A differenza dell'HCI, il fenomeno di BTI è parzialmente reversibile, poiché parte della carica viene liberata in seguito alla rimozione della tensione di gate. Infatti, in regime di funzionamento AC (cioè di tensione di gate variabile) l'aumento della tensione di soglia nel lungo periodo è inferiore rispetto a quello osservabile in regime di funzionamento DC (cioè di tensione di gate fissa, tale da mantenere il transistore pMOS acceso). Durante gli intervalli di tempo in cui il transistor è spento, si assiste ad un consistente ripristino della tensione di soglia, mentre, quando il transistor viene acceso, inizialmente la degradazione si ripresenta rapidamente, ma poi continua più lentamente. Complessivamente ciò risulta in un minore spostamento della tensione di soglia rispetto al caso in cui il transistor è continuamente sottoposto ad una tensione di gate-source negativa. Per questo motivo, in alcuni progetti low-power, la tradizionale tecnica di power saving nota come clock gating viene sostituita da un semplice abbassamento di frequenza, in modo da non spegnere completamente il clock e mitigare così gli effetti dell'NBTI.

Il modello fornito dal tool MOSRA per il BTI considera entrambi i fenomeni fisici, separandone i contributi in due diversi termini: uno determinato dalle trappole all'interfaccia e l'altro legato alle trappole che si formano all'interno del dielettrico [6, 7]. L'effetto di parziale ripristino delle condizioni del dispositivo in seguito ad una degradazione prestazionale è modellato considerando il duty cycle dello stress (regime di funzionamento AC) cui è sottoposto il dispostivo durante il suo normale funzionamento.

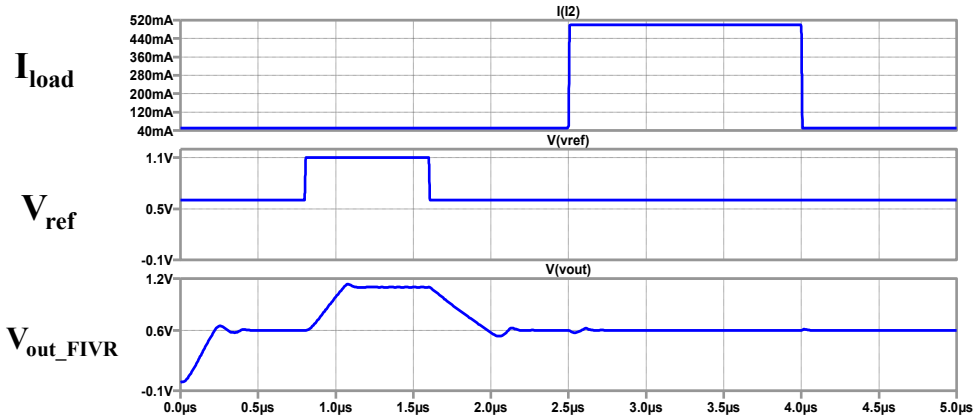


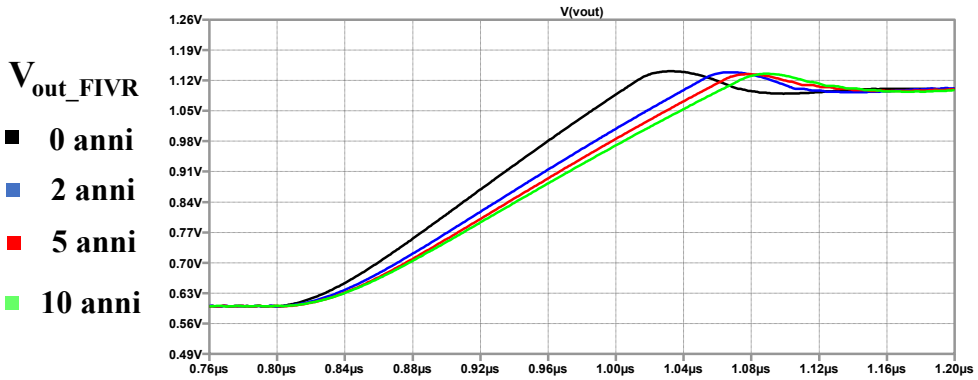
Figura 1.5: Simulazione in condizioni nominali. [5]

1.5 Effetti dell'aging sul FIVR

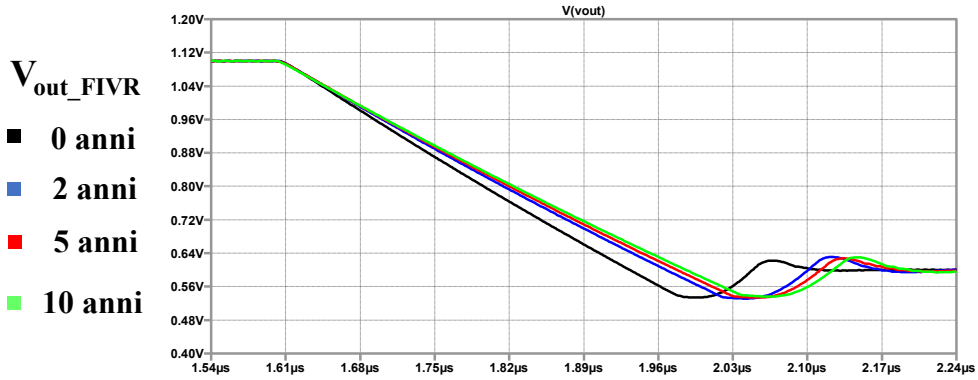
Vengono di seguito sinteticamente riportati alcuni risultati ottenuti dall'analisi presente in [5]. Per l'analisi è stato utilizzato il tool MOSRA sull'implementazione elettrica del FIVR in [4]. Il tool fornisce le tensioni di soglia dei transistori dopo un determinato tempo di lavoro per un particolare duty cycle richiesto. In tale analisi sono state effettuate le simulazioni per dieci anni di lavoro con duty cycle del 40%, 50%, 70% e 90%. Le quaranta tensioni di soglia ottenute sono state utilizzate da HSpice per valutare le prestazioni del FIVR, in termini di deterioramento delle forme d'onda.

In figura 1.5 vengono riportate le forme d'onda della tensione di riferimento V_{ref} e della corrente di carico I_{load} con il corrispondente andamento della tensione di uscita V_{out} fornita dal FIVR. Il riferimento è stato fatto variare tra i due estremi del possibile range di tensioni, cioè da 0.6 a $1.1V$ e viceversa. Si osserva come il FIVR impieghi del tempo per assestare l'uscita al valore di regime fornitogli dalla PCU. Successivamente viene simulato un cambiamento della corrente richiesta dal carico, passando da $50mA$ a $500mA$ e viceversa, per poter osservare come questo non influenzi la tensione di uscita del FIVR.

La parola di riferimento fornita dalla PCU è stata fatta variare in modo da ottenere le transizioni di caso peggiore, cioè quelle di salita e discesa della tensione di uscita ai due estremi del range possibile $[0.6, 1.1]V$, per valutare il degrado delle prestazioni determinato dall'aging. L'invecchiamento influisce in maniera evidente sul tempo impiegato dal nodo di uscita per assestarsi al valore di tensione corretto. In particolare, la maggior parte della degradazione si verifica nei primi due anni di attività, come si può osservare dalle figure 1.6.



(a) Transitorio di salita di V_{out} da 0.6 a $1.1V$.



(b) Transitorio di discesa di V_{out} da 1.1 a $0.6V$.

Figura 1.6: Simulazioni in presenza di aging con duty cycle medio del 50%. [5]

I dati presenti ricavati in [5] mostrano come l'aging deteriori notevolmente le durate dei transitori. Nonostante i valori di regime rimangano corretti, il calo delle prestazioni è evidente anche dopo pochi anni di utilizzo. In applicazioni caratterizzate da esigenze di elevata reliability, un rallentamento del circuito potrebbe non essere tollerabile, dunque si rende necessaria una strategia per rivelare gli effetti dell'aging sulle prestazioni del circuito.

1.6 Rivelatore proposto dallo standard ISO 26262

Lo standard ISO 26262 suggerisce una possibile implementazione del rivelatore per monitorare l'uscita del FIVR. Tale soluzione, schematizzata in figura 1.7, ricorre ad un convertitore analogico-digitale (ADC) per convertire in digitale (nove bits) la tensione fornita dal FIVR in modo

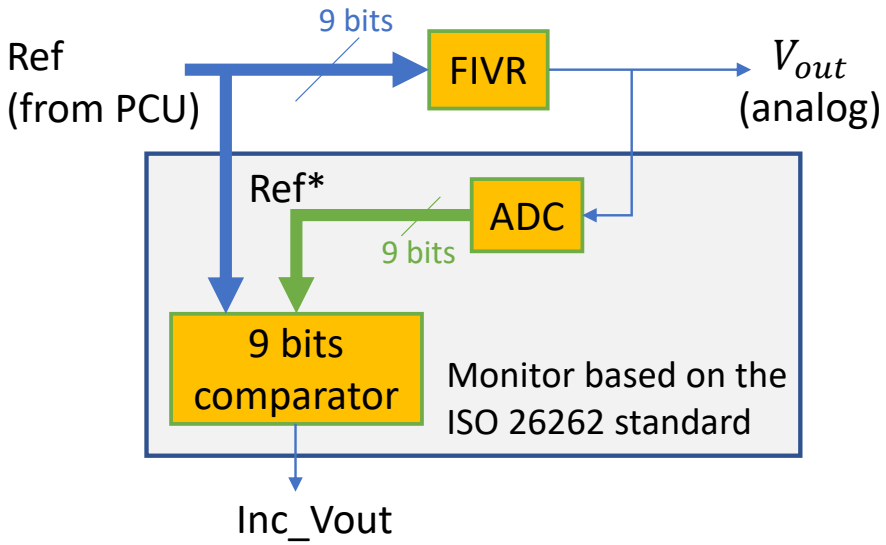


Figura 1.7: Schema della soluzione proposta dallo standard. [5, 11]

da poterla confrontare con la parola di riferimento (anch'essa di nove bits). Il comparatore ad alta affidabilità può essere ottenuto da un two-rail checker a nove bits (TRC_9), realizzato con otto TRC_2. Come ADC si è ipotizzato di utilizzare quello trovato in [10].

Pur essendo il funzionamento molto semplice in linea di principio, lo schema suggerito dallo standard non risulta adatto per la rivelazione degli eventuali guasti che possono verificarsi nel FIVR a causa dell'elevato costo. In particolare, affinché il monitor sia in grado di rivelare un'eventuale ripple con ampiezza eccessiva rispetto al margine di tolleranza, è necessario un ADC molto veloce, determinando così una grande occupazione d'area. Inoltre, non è a priori garantito che lo schema sia self-checking. Per queste ragioni, nel capitolo successivo viene proposto un rivelatore differente, caratterizzato da migliori prestazioni e da un costo nettamente inferiore, compatibile con l'applicazione cui è rivolto, come verrà mostrato nell'ultimo capitolo.

Capitolo 2

Strategia di rivelazione dei guasti del FIVR proposta

In questo capitolo viene presentata la nuova strategia per la rivelazione degli eventuali guasti che possono affliggere il FIVR. Sono stati definiti critici tutti quei guasti tali da portare la tensione di uscita del FIVR fuori da un margine di tolleranza prestabilito (in seguito definito). Lo schema di rivelazione proposto, che si basa su un'idea completamente differente rispetto a quella suggerita dallo standard ISO 26262, consente di rivelare guasti nel circuito durante il normale funzionamento sul campo.

L'approccio che era stato inizialmente proposto in [5] si basava sul confronto della tensione fornita dal FIVR con il riferimento rigenerato da un DAC ausiliario, analogo a quello già presente nel FIVR. In realtà, a partire dal riferimento digitale di nove bits fornito dalla PCU, si generavano due tensioni "analogiche" che costituivano il margine di tolleranza al di fuori del quale veniva segnalato il guasto. Tuttavia, risultava problematica la compensazione dell'offset dei due comparatori utilizzati per il confronto di V_{out} con le due soglie, nonché la precisione con cui venivano generate queste soglie.

2.1 Idea di base

La nuova strategia proposta valuta la differenza di tensione tra il riferimento prescritto dalla PCU e l'uscita del FIVR come differenza fra ritardi di propagazione. Come mostrato in figura 2.1, il FIVR riceve in ingresso la parola di nove bits che corrisponde al riferimento di tensione dato dalla PCU sulla base del monitoraggio dell'attività del power domain. La medesima parola di riferimento viene fornita anche al monitor per essere confrontata con la tensione di uscita del FIVR V_{out} secondo la strategia illustrata alla sezione successiva. Il monitor presenta due

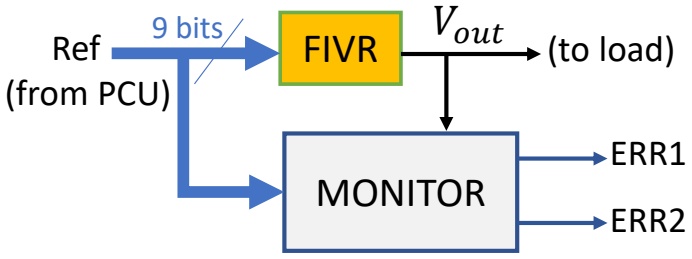


Figura 2.1: Schema ad alto livello della strategia proposta.

uscite (`ERR1,ERR2`) che forniscono un'indicazione circa il corretto funzionamento del circuito, comprensivo non solo del FIVR, ma anche del monitor stesso, visto che anche quest'ultimo può essere affetto da un guasto interno (per quanto questo sia un evento meno probabile rispetto ad un guasto al FIVR, in virtù delle minori complessità e area). La necessità che il monitor abbia due uscite, anziché una sola, deriva dal fatto che un banale SA della singola uscita potrebbe stare ad indicare sempre il corretto funzionamento del circuito, aspetto particolarmente critico in contesti di alta affidabilità.

La strategia di rivelazione proposta permette non solo di ottenere un'indicazione d'errore qualora la tensione di uscita del FIVR esca dal margine di tolleranza (pari a 1 LSB della parola digitale di riferimento a nove bits data dalla PCU) per un guasto interno al FIVR, ma consente anche di rivelare il rallentamento dei transitori del FIVR dovuto a fenomeni di invecchiamento dei dispositivi. Tale schema è in grado di rivelare anche del ripple con ampiezza eccedente il margine di tolleranza sulla tensione di uscita, nonostante il valore medio possa essere corretto.

Una possibile implementazione a livello elettrico verrà illustrata nel capitolo seguente.

2.2 Schema a blocchi

In figura 2.2 viene mostrato lo schema a blocchi del monitor progettato in questa tesi. Il riferimento digitale di nove bits proveniente dalla PCU, che viene fornito al FIVR per generare la tensione V_{out}, viene dato anche ad un DAC ausiliario (analogo a quello presente nel FIVR) interno al monitor per rigenerare il riferimento come tensione "analogica" V_{ref}. Ciascuna delle due voltage-controlled delay lines (VCDLs), costituite da una cascata di celle di ritardo elementari, viene controllata dalla corrispondente tensione di controllo, che ne regola il ritardo. Alle due linee di ritardo viene dato in ingresso il medesimo segnale di clock, che si propaga lungo ciascuna delle due VCDLs con un ritardo dipendente dalla

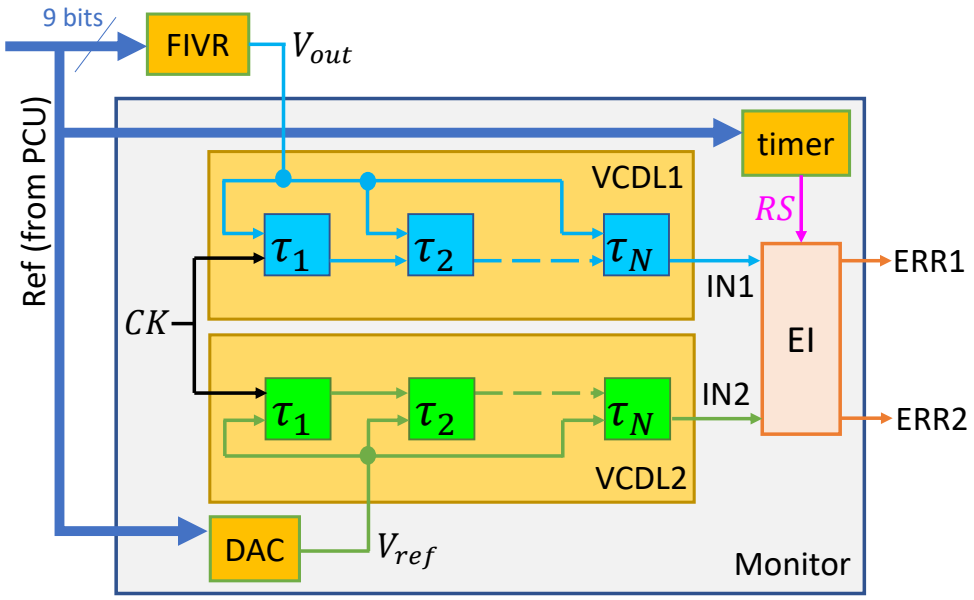


Figura 2.2: Schema a blocchi del monitor proposto.

corrispondente tensione di controllo. In questo modo la differenza (assoluta) di tensione $|V_{ref} - V_{out}|$ viene tradotta in una differenza (assoluta) di ritardi $|d_1 - d_2|$. Le uscite delle VCDLs vengono date in ingresso ad un error indicator (EI), che genera e memorizza un'indicazione d'errore qualora la differenza di ritardi $|d_1 - d_2|$ sia superiore ad una certa soglia temporale, corrispondente ad una differenza di tensioni $|V_{ref} - V_{out}|$ pari a 1 LSB, di ampiezza pari a $V_{in}/2^9 \simeq 3.5mV$.

Le celle di ritardo della linea `VCDL1` (`VCDL2`) sono state implementate in modo tale che il loro ritardo sia proporzionale alla tensione di controllo V_{out} (V_{ref}). In questo modo il ritardo di propagazione d_1 (d_2) dei fronti di salita e discesa del segnale di clock attraverso la linea `VCDL1` (`VCDL2`) dall'ingresso `CK` all'uscita `IN1` (`IN2`) è proporzionale alla tensione V_{out} (V_{ref}). Da questa implementazione segue la proporzionalità $|V_{ref} - V_{out}| \propto |d_1 - d_2|$, alla base del funzionamento del monitor.

In presenza di ingressi `(IN1,IN2)` con valori logici complementari (cioè di segnali non allineati per il diverso ritardo dovuto ad una certa differenza di tensione), l'EI memorizza alle proprie uscite `(ERR1,ERR2)` un'indicazione d'errore, che viene mantenuta finché non viene applicato un segnale di reset (`RS`). La memorizzazione dell'indicazione d'errore risulta fondamentale, dato che la sua durata sarebbe solo temporanea altrimenti.

In caso di tensione di uscita del FIVR V_{out} corretta (cioè valori di V_{out} all'interno del margine di tolleranza di 1 LSB rispetto al riferimento), la

IN1	IN2	Significato
1	1	Ok
1	0	Errore
0	1	Errore
0	0	Ok

Tabella 2.1: Codifica dell'EI.

differenza tra i ritardi dei segnali in uscita dalle due VCDLs è minore della sensibilità dell'EI (by design). In questo caso l'EI interpreta gli ingressi `(IN1,IN2)` come allo stesso valore logico (che si alterna alla frequenza del clock). L'EI è progettato in modo che anche le proprie uscite `(ERR1,ERR2)` assumano lo stesso valore logico (complementato) che si alterna alla frequenza del clock.

Il fatto che, nel caso fault-free, vi siano due possibili configurazioni che si alternano alle uscite dell'EI è particolarmente importante nella progettazione di circuiti self-checking (SCCs), cioè circuiti in grado di fornire un'indicazione circa la correttezza dell'uscita durante il normale funzionamento sul campo tramite una sorta di continuo autocollaudo. Infatti, facendo le convenzionali ipotesi di guasto singolo (cioè i guasti si verificano uno alla volta) e di tempo sufficientemente lungo tra la comparsa di due guasti, se una delle due uscite dell'EI dovesse bloccarsi ad un determinato valore logico (SA), l'altra uscita continuerebbe a funzionare normalmente, portando alla segnalazione dell'errore al primo cambio di valore.

In caso di tensione di uscita del FIVR V_{out} errata (cioè valori di V_{out} al di fuori del margine di tolleranza di 1 LSB rispetto al riferimento), la differenza tra i ritardi dei segnali in uscita dalle due VCDLs è maggiore della sensibilità dell'EI. In questo caso l'EI rileva che per un certo intervallo di tempo gli ingressi `(IN1,IN2)` assumono valori logici complementari. L'EI è progettato in modo da memorizzare un'indicazione d'errore alle proprie uscite finché non viene applicato un segnale di reset (`RS=1`). La codifica delle parole di codice e delle indicazioni d'errore è mostrata in tabella 2.1.

Il segnale di reset `RS` viene generato da un blocco denominato "timer" in modo da evitare i falsi positivi, cioè le indicazioni d'errore errate durante il transitorio di risposta del FIVR (t_R) a seguito di un cambio del riferimento, in cui la tensione V_{out} potrebbe essere al di fuori del margine di tolleranza. La durata del reset è determinata dal massimo tempo consentito per il transitorio del FIVR ($t_{R_{max}}$), che è stato calcolato come il tempo che impiega il FIVR per portarsi entro il margine di

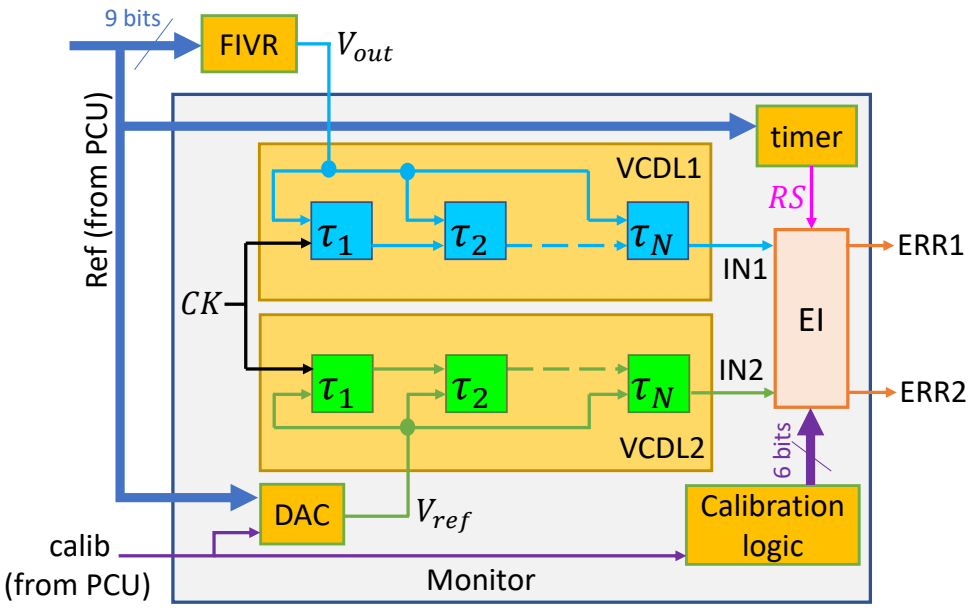


Figura 2.3: Schema a blocchi completo del monitor proposto.

tolleranza nella transizione di caso peggiore. La transizione di caso peggiore è stata individuata essere quella di discesa dalla tensione massima di $1.1V$ a quella minima di $0.6V$. Dalle simulazioni effettuate sull'implementazione del FIVR fatta in [4] risulta che tale transizione dura $t_{R_{max}} \simeq 502ns$. Quando `RS=1`, l'EI viene inibito, per cui le sue uscite si portano a `(ERR1,ERR2)=(0,0)`. Mentre, negli intervalli di tempo in cui la tensione V_{out} dovrebbe essere stabile, cioè dopo il massimo intervallo di tempo consentito per concludere il transitorio $t_{R_{max}}$, in presenza di una tensione di uscita corretta, si ha `(ERR1,ERR2)=(`$\overline{\texttt{IN1}}$`,`$\overline{\texttt{IN2}}$`)`.

2.3 Strategia di calibrazione del monitor

Per rendere lo schema di rivelazione robusto alle *process parameters variations* (PPVs), è necessario fornire la possibilità di calibrare la sensibilità dell'EI. Infatti, le PPVs che affliggono sia l'EI sia le VCDLs incidono profondamente sull'efficacia dello schema di rivelazione, perché la soglia che permette di discriminare fra circuito guasto o correttamente funzionante viene alterata. Dunque, è possibile che tale soglia non consenta di segnalare un errore in presenza di un guasto. Questa costituisce una situazione potenzialmente critica, perché in alta affidabilità non è tollerabile avere una falsa indicazione di corretto funzionamento del sistema. Risulta, quindi, fondamentale calibrare la sensibilità dell'EI dopo la fabbricazione in modo da garantire il margine di tolleranza imposto

durante la fase di progettazione (in questo caso pari a 1 LSB) per ogni chip fabbricato. In questo modo l'efficacia del monitor non viene compromessa dalle PPVs. Quindi, allo schema precedentemente mostrato in figura 2.2, è necessario aggiungere un blocco che si occupa di effettuare la calibrazione. Lo schema completo viene riportato in figura 2.3.

Come verrà spiegato più nel dettaglio alla sezione 3.3.1 del capitolo successivo, si può pensare di effettuare la calibrazione non solo a seguito della fabbricazione, ma anche ad ogni accensione del sistema, in modo da compensare pure eventuali variazioni indotte dall'aging e dalla temperatura cui opera il circuito.

Per la calibrazione si assume di poter avere a disposizione un apposito segnale di un bit fornito dalla PCU ad ogni accensione del sistema. Questo segnale, denominato `calib`, viene fornito sia al DAC interno al monitor sia ad un blocco che si occupa di generare i sei bits di calibrazione dati all'EI per effettuare il tuning. Occorre fornire tale segnale anche al DAC in quanto, per ottenere il margine di tolleranza desiderato, è necessario che V_{ref} e V_{out} differiscano di 1 LSB durante la fase di calibrazione.

Capitolo 3

Possibile implementazione a livello elettrico del monitor

In questo capitolo, facendo riferimento alla figura 2.2, viene mostrata una possibile implementazione a livello elettrico del monitor basato sulla strategia di rivelazione proposta nel capitolo precedente. Le simulazioni sono state effettuate con LTSpice, utilizzando il modello predittivo di tecnologia a $22nm$ fornito dall'Arizona State University [12].

L'intero monitor è alimentato dalla tensione $V_{in}/2 = 0.9V$, disponibile all'interno del FIVR come riferimento stabile, ottenuto da un regolatore low-dropout (LDO) a partire dalla tensione V_{in}. I vantaggi di un regolatore LDO rispetto ad un tradizionale convertitore DC/DC sono dati dall'assenza di fenomeni di switching e da una minore occupazione d'area, poiché non sono necessari elementi per il filtraggio della tensione di uscita.

Per l'implementazione a livello elettrico delle varie porte logiche sono stati tenuti in considerazione (laddove possibile) alcuni accorgimenti per risparmiare potenza assorbita dall'alimentazione e ridurre i tempi di propagazione [13]. In particolare, collegare i segnali con minore probabilità di transizione ai transistori con source a potenziale costante implica che le capacità parassite dei nodi interni vengono caricate e scaricate meno frequentemente, determinando un minore consumo di potenza dinamica. Incidentalmente, questo risulta più vantaggioso anche in termini di tempi di propagazione della porta logica, poiché la commutazione del segnale collegato al gate di un transistore con source a potenziale costante richiede di caricare/scaricare non solo la capacità di uscita, ma anche quella del nodo interno. Inoltre, per velocizzare i tempi di propagazione delle porte logiche è possibile collegare i segnali più lenti (cioè quelli le cui commutazioni sono ritardate rispetto agli altri segnali in ingresso alla medesima porta logica) ai transistori più prossimi al no-

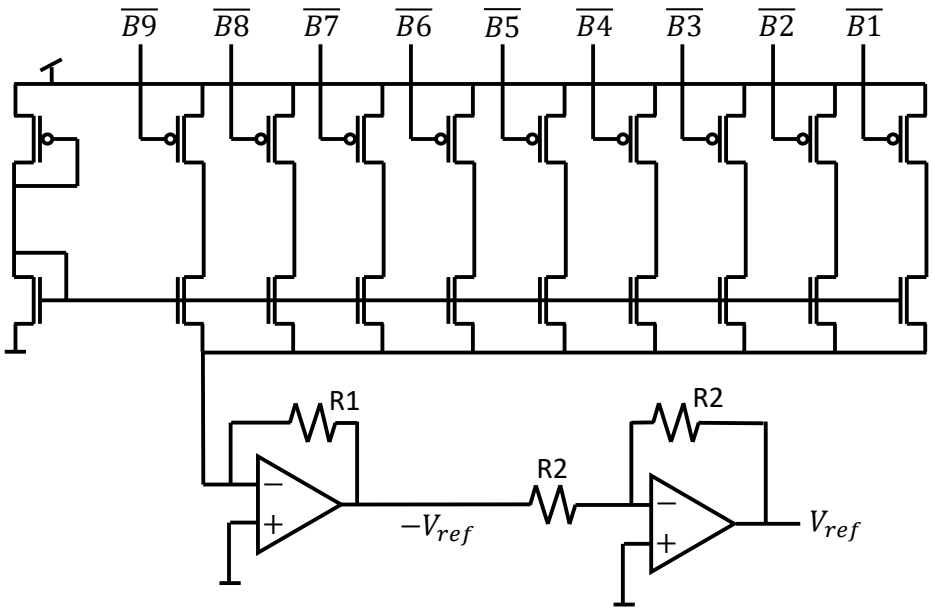


Figura 3.1: Schema del DAC interno al monitor.

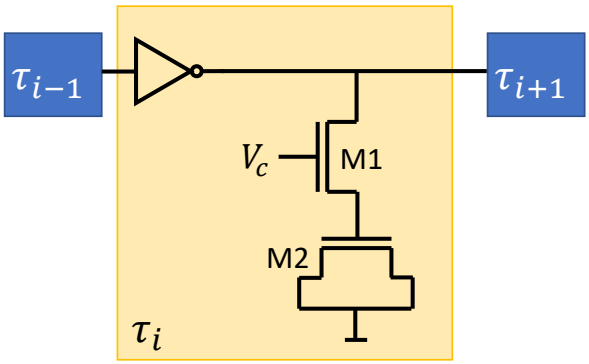
do di uscita in modo che, quando tali segnali commutano, deve essere caricata/scaricata solo la capacità di uscita, poiché le capacità nei nodi interni sono già state caricate/scaricate dagli altri segnali, che hanno commutato prima, essendo più veloci.

3.1 DAC

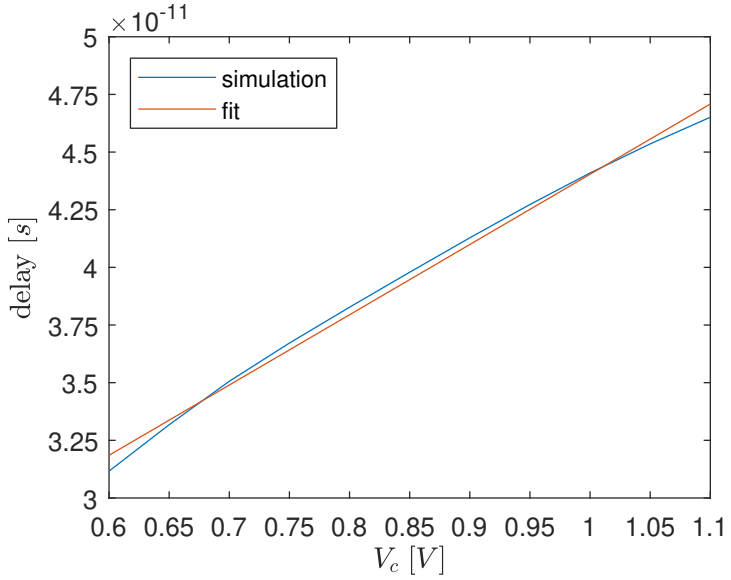
Come implementazione del DAC si è mantenuta quella utilizzata in [4], che riprende quella proposta in [14], e viene mostrata in figura 3.1. Il DAC è dello stesso tipo di quello già presente nel FIVR per convertire il riferimento digitale di nove bits nella tensione "analogica" V_{ref}. Il principio di funzionamento è molto semplice: ogni ramo genera una corrente proporzionale a quanto è significativo il bit che lo pilota, cioè ciascun ramo eroga una corrente doppia rispetto a quello alla propria destra. Le correnti di tutti i rami si sommano nel nodo di ingresso di un convertitore I/V, costituito da un OpAmp con la resistenza `R1` in retroazione. L'uscita di tale convertitore è la tensione $-V_{ref}$, che viene cambiata di segno da un amplificatore invertente a guadagno unitario.

3.2 Voltage Controlled Delay Lines

Le VCDLs sono costituite da una cascata di celle di ritardo elementari, ciascuna delle quali viene controllata da una tensione analogica (V_{out} nel



(a) Implementazione della cella di ritardo.



(b) Caratteristica della cella di ritardo.

Figura 3.2

caso della `VCDL1` o V_{ref} nel caso della `VCDL2`). Per scegliere la topologia circuitale della singola cella, sono state esaminate e simulate diverse implementazioni, riportate in [15–18]. In generale, gli elementi di ritardo controllati da una tensione "analogica" (cioè una tensione che può assumere un determinato valore all'interno di un range prestabilito) sono basati sull'utilizzo di specchi di corrente, in cui è presente un transistore saturo la cui tensione di gate regola la corrente di drain con una dipendenza approssimativamente lineare (per le tecnologie scalate, il modello quadratico del MOSFET in saturazione fornisce una forte sovrastima della corrente). Tuttavia, le topologie analizzate godevano di una discreta linearità tra tensione di controllo e ritardo solo in un range molto

Caso	Ritardo $[ps]$
Nominale	1.20
Massimo	2.00
Minimo	0.80

Tabella 3.1: Risultato delle simulazioni Monte Carlo sulle VCDLs.

ristretto di tensioni, soprattutto limitato superiormente dalla tensione di alimentazione pari a $V_{in}/2 = 0.9V$.

L'unica implementazione che garantisse una buona linearità in tutto il range di possibile variazione della tensione di uscita del FIVR (da 0.6 a $1.1V$) è quella mostrata in figura 3.2a, costituta da un inverter, seguito dal condensatore $M2$ (realizzato con un nMOS con i terminali di source e drain corto-circuitati) che viene caricato dalla corrente che fluisce nel transistore $M1$, la cui conducibilità viene controllata dalla tensione V_c. Da un punto di vista qualitativo, all'aumentare della tensione V_c aumenta la corrente che scorre in $M1$, che va a sottrarsi alla corrente erogata dall'inverter, riducendo quella disponibile per la carica della capacità di gate della porta logica a valle (cioè l'inverter della cella di memoria successiva). Questo fa sì che il ritardo aumenti.

La risoluzione del monitor è proporzionale alla lunghezza delle linee di ritardo (cioè al numero di celle connesse in cascata) e alla pendenza della caratteristica ritardo-vs-V_c della singola cella di ritardo. Il dimensionamento degli elementi di ritardo è stato ottenuto effettuando alcune simulazioni con l'obiettivo di ottimizzare il compromesso tra la pendenza della caratteristica e l'area occupata. Come si può notare dalla figura 3.2b, la caratteristica presenta una buona linearità in tutto il range di tensioni di uscita V_{out}. Il ritardo è stato misurato al 50% dell'escursione del segnale tra fronti omologhi distanti due celle di ritardo, in modo da compensare il diverso ritardo tra un fronte di salita ed uno di discesa.

Per valutare la robustezza dello schema rispetto alle PPVs, sono state effettuate delle simulazioni Monte Carlo in modo da individuare le possibili variazioni nella differenza dei ritardi tra le due linee. Per le simulazioni sono stati variati in maniera indipendente spessore dell'ossido, mobilità e tensione di soglia secondo distribuzioni uniformi con variazioni fino al 20% rispetto al valore nominale. I risultati sono riportati in tabella 3.1.

3.3 Error Indicator

L'EI si basa sull'implementazione presente in [19] con alcune modifiche dovute all'inversione tra la codifica delle parole di codice ed indicazioni

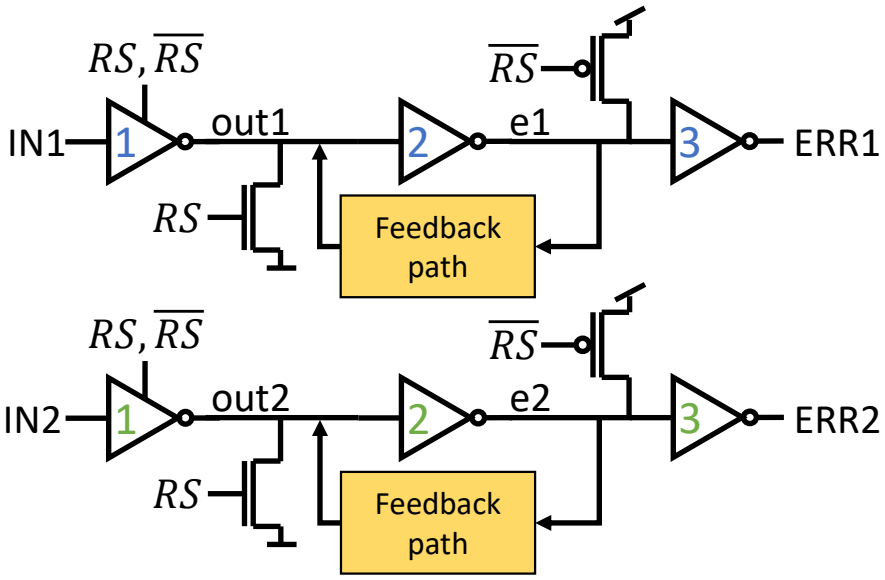


Figura 3.3: Rappresentazione schematica dell'EI.

d'errore e l'introduzione della possibilità di tuning per regolare la sensibilità dell'EI, in modo tale da rendere il monitor robusto alle PPVs. Lo schema a blocchi è mostrato in figura 3.3.

Durante il normale funzionamento del circuito `RS=0`, quindi i buffers tri-state denominati come `NOT1` sono conduttivi. Nel caso fault-free `IN1=IN2` (cioè la differenza tra i ritardi è nulla), dunque agli ingressi dell'EI appaiono le configurazioni `(0,0)` o `(1,1)` che si alternano alla frequenza del clock. In caso di guasto, i segnali `IN1` e `IN2` non sono allineati (cioè la differenza tra i ritardi è maggiore di zero), dunque la configurazione `(0,1)` o `(1,0)` (la cui durata è proporzionale alla differenza di tensioni $|V_{ref} - V_{out}|$) appare temporaneamente agli ingressi dell'EI. Se la durata di tale configurazione è sufficientemente lunga, allora il cammino di feedback (invertente) diventa conduttivo e memorizza l'indicazione d'errore, essendo dimensionato per essere dominante rispetto ai buffers tri-state `NOT1`. L'indicazione d'errore rimane memorizzata fino all'applicazione del segnale di reset (`RS=1`). Il funzionamento dell'EI è riassunto in tabella 3.2.

Durante la fase di reset i buffers tri-state `NOT1` vengono interdetti, quindi le corrispettive uscite `out1` e `out2` si portano in uno stato di alta impedenza. I transistori nMOS pilotati dal segnale `RS` forzano uno zero sui nodi `out1` e `out2`, mentre i pMOS pilotati dal segnale $\overline{\texttt{RS}}$ forzano un valore logico alto sui nodi `e1` e `e2`. In questo modo le uscite dell'EI si portano a `(ERR1,ERR2)=(0,0)`. Teoricamente, una volta che i buffers tri-state sono passati nello stato di alta impedenza, sarebbero

e1	e2	Feedback paths	out1	out2
0	0	Alta impedenza	$\overline{\texttt{IN1}}$	$\overline{\texttt{IN2}}$
0	1	Conduttivi	$\overline{\texttt{e1}}$	$\overline{\texttt{e2}}$
1	0	Conduttivi	$\overline{\texttt{e1}}$	$\overline{\texttt{e2}}$
1	1	Alta impedenza	$\overline{\texttt{IN1}}$	$\overline{\texttt{IN2}}$

Tabella 3.2: Funzionamento dell'EI.

sufficienti gli nMOS per effettuare il reset dell'EI. Tuttavia, per rendere gli nMOS dominanti rispetto ai cammini di feedback, sarebbe necessaria una larghezza di canale eccessivamente grande, che introdurrebbe un elevato carico capacitivo sui nodi `out1` e `out2`, rallentando i transitori dei buffers `NOT1` durante il normale funzionamento dell'EI.

Per ovviare al problema delle PPVs, che agiscono sia sull'EI sia sulle VCDLs, è stata introdotta la possibilità di effettuare il tuning dell'EI in maniera tale da mantenere inalterato il margine di tolleranza di 1 LSB. Il tuning avviene controllando, in maniera digitale, la conducibilità dei cammini di feedback, grazie alla presenza di un ulteriore ramo di pull-down e del suo complementare pull-up, come mostrato in figura 3.4. Cinque bits $\texttt{t}_\texttt{i}$ (ed i relativi complementi $\overline{\texttt{t}_\texttt{i}}$) permettono di accendere o spegnere tali transistori secondo una codifica binaria che permette di variare la conducibilità in una scala da 0 (in realtà almeno un transistore deve essere acceso affinché il cammino di feedback sia conduttivo) a $2^5 - 1 = 31$. Anziché aumentare progressivamente la larghezza di canale dal LSB al MSB (cioè da $\texttt{t}_\texttt{1}$ a $\texttt{t}_\texttt{5}$), si è deciso di aumentare il numero di *fingers* che costituiscono il transistore tramite il parametro m. Per il tuning è presente un bit addizionale, denominato `ENfb`, necessario per aumentare la conducibilità dei transistori pilotati dai nodi `out1` e `out2` in casi particolari. Tale bit viene trattato come se fosse più significativo rispetto a $\texttt{t}_\texttt{5}$.

Il dimensionamento del circuito è stato effettuato grazie a varie simulazioni parametriche. In particolare si è dimensionato il circuito in modo tale che, in condizioni nominali sia dell'EI sia delle VCDLs, l'EI fosse in grado di segnalare l'errore per una differenza di tensione pari a 1 LSB con la metà dei bits accesi, cioè $t_i = 1\ i = 1, \ldots, 4$. Successivamente sono state effettuate delle simulazioni Monte Carlo, i cui risultati sono stati elaborati con uno script Matlab creato ad-hoc, per individuare i casi opposti di massima e minima sensibilità, cioè rispettivamente di minima e massima tolleranza. Questi due casi estremi di PPVs dell'EI sono stati confrontati con i casi opposti di massima e minima differenze di ritardi tra i segnali in uscita dalle due linee di ritardo. In particolare:

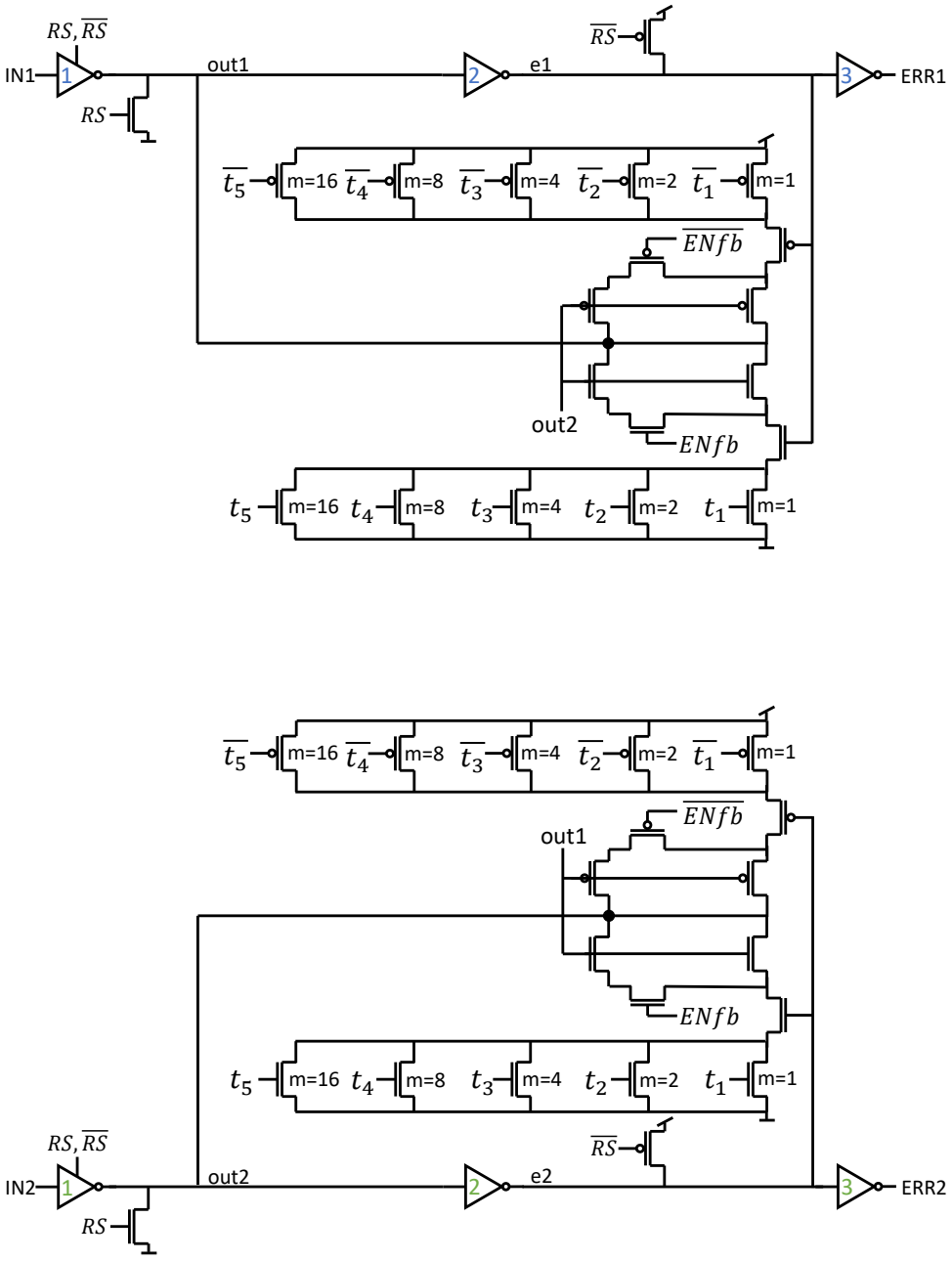


Figura 3.4: Schema elettrico dell'EI.

- il caso di minima sensibilità (cioè di massima tolleranza) dell'EI è stato confrontato con il caso di minima differenza tra i ritardi delle VCDLs: questa è la situazione più critica per la safety, perché un possibile guasto viene rivelato solo per un'elevata differenza tra V_{out} e V_{ref}, dunque si è agito sui bits di tuning dell'EI in modo da

Ritardo VCDLs [ps]	Sensibilità EI	Bits on	Sensibilità EI ottenuta [ps]	Margine di tolleranza ottenuto [# LSB]
0.80	Minima	$\mathtt{t_5}$,ENfb	0.90	$0.90p/0.80p = 1.125$
2.00	Massima	$\mathtt{t_1}$,$\mathtt{t_2}$	2.90	$2.90p/2.00p = 1.45$

Tabella 3.3: Risultati del confronto tra VCDLs e EI in caso di PPVs.

aumentarne la sensibilità

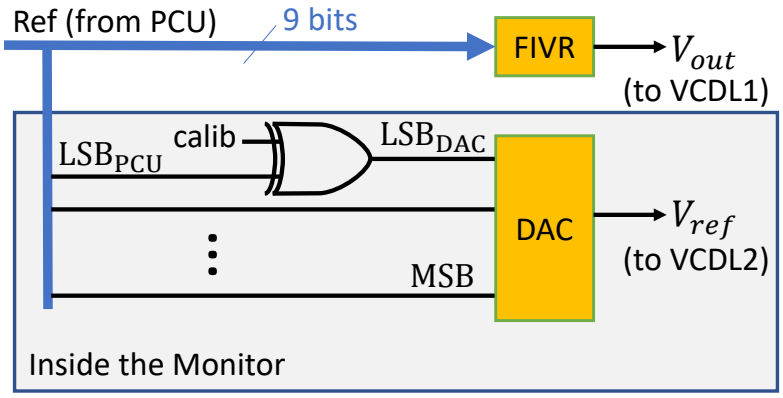
- il caso di massima sensibilità (cioè di minima tolleranza) dell'EI è stato confrontato con il caso di massima differenza tra i ritardi delle VCDLs: in realtà, questa non costituisce una criticità per la safety, dato che il sistema risulta molto meno tollerante rispetto al caso nominale, tuttavia, allo scopo di evitare la segnalazione di falsi positivi, si è agito sui bits di tuning dell'EI in modo da ridurne la sensibilità

Anche in questo caso, per le simulazioni Monte Carlo si è assunto che i parametri variassero con distribuzioni uniformi con variazioni fino al 20% rispetto ai rispettivi valori nominali. Inoltre, si è assunto che le PPVs cui è soggetto l'EI siano indipendenti da quelle che affliggono le linee di ritardo. I risultati del confronto di cui sopra sono riportati in tabella 3.3. Si può osservare come, nel caso peggiore tra le due situazioni estreme considerate, si riesca ad ottenere un margine di tolleranza di soli 1.45LSB, corrispondenti a circa $1.45 \times 3.5m = 5.075mV$.

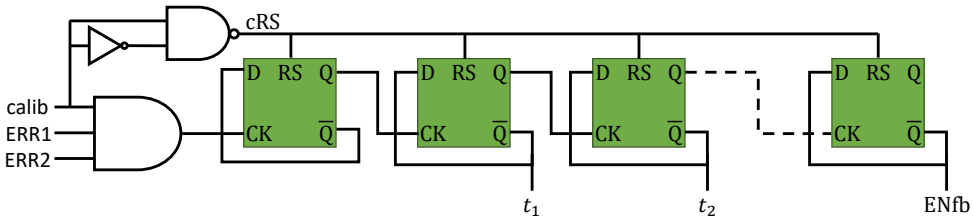
3.3.1 Circuito di calibrazione dell'EI

Per calibrare l'EI mostrato in figura 3.4 in modo da compensare le PPVs e garantire un margine di tolleranza di 1 LSB, è necessario introdurre dell'hardware aggiuntivo che opera la calibrazione ad ogni accensione del sistema. La scelta di effettuare tale calibrazione ad ogni accensione del sistema deriva da due motivi: innanzitutto, diventa così possibile la compensazione rispetto a fenomeni di aging che possono affliggere anche l'EI e variazioni di temperatura, garantendo in qualsiasi caso il margine di tolleranza desiderato, inoltre, se la calibrazione venisse fatta solo una tantum dopo la fabbricazione, sarebbe necessario aggiungere una memoria in modo da conservare il valore dei bits $\mathtt{t_i}$ ($i = 1, \ldots, 5$) e ENfb.

L'hardware aggiunto è mostrato in figura 3.5. Il circuito si basa su un contatore binario a sei bits ($\mathtt{t_i}$ con l'aggiunta di ENfb, cioè dal LSB al MSB), come mostrato in figura 3.5b, che comincia a contare da 0 e può (eventualmente) arrivare a $2^6 - 1 = 63$. Ogni bit del contatore è ottenuto a partire da un semplice divisore di frequenza per due, imple-



(a) Modifica al DAC ausiliario.



(b) Contatore binario.

Figura 3.5: Circuito di calibrazione dell'EI.

mentato come un D-flip-flop (FF) in cui l'uscita negata $\overline{\texttt{Q}}$ è retroazionata sull'ingresso di dato `D`: in questo modo il segnale $\overline{\texttt{Q}}$ (e di conseguenza anche `Q`) possiede una frequenza dimezzata rispetto a quella del segnale dato all'ingresso di clock `CK`. Ogni FF può essere resettato grazie alla presenza di un NAND gate interno a ciascun latch che costituisce il FF: con `cRS=0` tutti i FFs portano la rispettiva uscita a `Q=1`. Un ulteriore FF viene inserito a monte del contatore in modo da rallentare ulteriormente il LSB $\texttt{t}_1$ (e di conseguenza i bits più significativi seguenti) rispetto alla frequenza dei segnali `ERR1` ed `ERR2`, in modo che l'EI abbia tempo a sufficienza per poter eventualmente rivelare l'errore. Il contatore è asincrono, perché la commutazione dei bits non è data da un unico segnale di clock condiviso tra i FFs.

Come visto precedentemente, la parola di riferimento di nove bits fornita dalla PCU al FIVR viene data anche al DAC ausiliario interno al monitor per rigenerare il riferimento V_{ref}. Durante il processo di calibrazione si ha `calib=1`, quindi l'EXOR gate in figura 3.5a inverte il LSB della parola generata dalla PCU e lo dà in ingresso al DAC, cioè $\texttt{LSB}_{\texttt{DAC}}=\overline{\texttt{LSB}_{\texttt{PCU}}}$. In questo modo, durante la calibrazione, V_{out} e V_{ref} differiscono di 1 LSB, mentre, nel normale funzionamento del monitor, si ha `calib=0`, dunque il LSB non viene alterato.

Da notare che, sotto l'ipotesi che la parola fornita dalla PCU durante la fase di calibrazione sia fissata a priori, è possibile scegliere una parola con $\texttt{LSB}_{\texttt{PCU}}\texttt{=0}$ in modo che, sostituendo l'EXOR gate con una porta OR, questo dia $\texttt{LSB}_{\texttt{DAC}}\texttt{=1}$, semplificando ulteriormente il circuito. La semplificazione deriva dal fatto che per implementare una porta EXOR sono richiesti due NOT aggiuntivi, poiché l'EXOR richiede di avere gli ingressi sia in foma vera sia in forma negata, mentre per implementare una porta OR sono sufficienti un NOR ed un NOT. Questa semplificazione permette di ottenere un lieve risparmio in termini di area e potenza dissipata sul piano teorico (seppure abbondantemente trascurabili rispetto al totale del monitor dal punto di vista pratico).

Dopo l'accensione del sistema, `calib=1` per iniziare il processo di calibrazione, quindi il NAND gate in figura 3.5b ha un breve impulso negativo sulla propria uscita: `cRS=0` forza `Q=1` su tutti i FFs in modo che tutti i bits di tuning siano spenti, quindi i cammini di feedback interni all'EI non sono conduttivi. Poiché l'EI non è in grado di rivelare alcun guasto (dato che i cammini di feedback sono inibiti), le uscite dell'EI `(ERR1,ERR2)` assumono valori logici uguali che si alternano alla frequenza del segnale di clock dato in ingresso alle due VCDLs. Mentre `calib=1` l'AND gate è abilitato e fornisce un segnale di clock al primo FF, quindi il circuito comincia a contare. Quando l'EI fornisce l'indicazione d'errore (che nel caso fault-free è garantita esserci, poiché volutamente durante la calibrazione si ha che V_{ref} e V_{out} differiscono di 1 LSB come visto precedentemente), `ERR1` e `ERR2` assumono valori logici complementari e l'EI memorizza l'indicazione d'errore $(1, 0)$ o $(0, 1)$, dunque l'AND gate dà uno 0 in uscita e il contatore viene arrestato.

È necessario che il segnale di calibrazione `calib` rimanga attivo per il tempo massimo richiesto dal processo di calibrazione, che corrisponde al caso in cui il contatore deve contare fino alla configurazione finale corrispondente a $2^6 - 1 = 63$. Trascorso questo tempo, il segnale di calibrazione può essere disattivato (`calib=0`), quindi l'AND gate rimane inibito e nessun segnale di clock viene fornito al primo FF (una sorta di clock gating). In questo modo il contatore mantiene memorizzata la configurazione dei bits di tuning che permette all'EI di avere una sensibilità corrispondente ad 1 LSB. Dunque, il segnale di calibrazione rimane attivo per $T_{CK} \cdot 2^7$, poiché sette sono i FFs connessi in cascata per dividere progressivamente la frequenza per due.

Il circuito risulta self-checking in quanto, in caso di guasto interno, il contatore continua a funzionare correttamente, oppure il guasto interno viene rivelato dalla mancata indicazione d'errore dopo che è trascorso il tempo massimo consentito al processo di calibrazione, cioè dopo che il segnale `calib` ha commutato da 1 a 0.

3.4 Timer

Il timer è stato implementato come un tradizionale multivibratore monostabile basato su porte logiche, come mostrato in figura 3.6. Tale circuito viene utilizzato per effettuare il reset dell'EI sia nel caso di cambio della tensione di riferimento, in cui è necessario inibire l'EI onde evitare la segnalazione di falsi positivi, sia nel caso di reset di sistema per rimuovere un'indicazione d'errore a seguito dell'attivazione di un'opportuna procedura di *recovery*. L'aggiunta di due inverters prima dell'uscita è necessaria per ottenere una forma d'onda rigenerata.

Per quanto riguarda la generazione del segnale di reset nel caso di transitorio del FIVR, si è ipotizzato di avere a disposizione un segnale di 1 bit generato dalla PCU per campionare i nove bits della parola di riferimento in un registro all'ingresso del FIVR. In caso il segnale `sample_ref` presenti un impulso (che segnala il cambio della tensione di riferimento), il monostabile genera un impulso `RS=1` di durata pari a $T = RC \ln 2$ (nel caso ideale di porte logiche simmetriche) che viene imposto uguale a $t_{R_{max}}$. Per la generazione del segnale complementare $\overline{\texttt{RS}}$ si utilizza una replica del monostabile (aggiungendo un inverter all'uscita) in modo da evitare che guasti che possono affliggere il monostabile (ad esempio, un banale SA1 sull'uscita `RS`) vadano ad impattare l'efficienza nella rivelazione dei guasti nel FIVR da parte del monitor proposto in questa tesi.

Per determinare la durata dell'impulso generato dal monostabile, pari a $t_{R_{max}}$, è stata simulata la transizione di caso peggiore (cioè quella di discesa da $V_{out} = 0.6V$ a $1.1V$) sull'implementazione del FIVR fatta in [4]. Misurando il tempo che intercorre tra il cambio di riferimento e l'istante in cui V_{out} si porta entro un margine di tolleranza di 1 LSB dalla tensione di riferimento, si è trovato che $t_{R_{max}} \simeq 502ns$. Posto $C = 1pF$, nel caso ideale si ottiene una resistenza di valore pari a $R =$

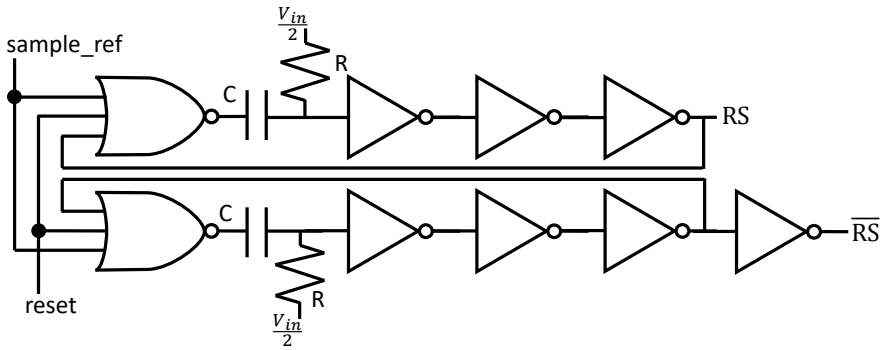


Figura 3.6: Schema del timer.

$502n/(\ln 2 \cdot 1p) \simeq 724k\Omega$. Poiché il monostabile è implementato con porte logiche reali (seppur dimensionate in modo da essere simmetriche), da simulazione si è trovato un valore di resistenza pari a $735k\Omega$.

3.4.1 Implementazione alternativa del monostabile

L'implementazione del timer come un tradizionale monostabile basato su porte logiche, seppure perfettamente funzionante da un punto di vista teorico, soffre di alcune problematiche che lo rendono inadatto ad un reale utilizzo.

Il problema maggiore è sicuramente costituito dalle PPVs che hanno una maggiore incidenza sui valori di capacità e resistenza fabbricati, oltre che sui transistori che compongono le porte logiche. Valori errati di resistenza e capacità modificano la costante di tempo del circuito variando la durata dell'impulso generato dal monostabile. In particolare, se le PPVs sono tali da aumentare tale durata, allora il monitor perde di efficacia nella rivelazione di eventuali rallentamenti nei transitori del FIVR dovuti a fenomeni di aging. Invece, il caso opposto di PPVs tali da ridurre la durata dell'impulso non rappresenterebbe un problema, perché il monitor risulterebbe più severo. Dunque, sarebbe di fondamentale importanza poter effettuare il tuning dopo la fabbricazione per compensare le PPVs, ma la tradizionale implementazione del monostabile non fornisce tale possibilità. Si potrebbe pensare di sostituire la resistenza con un amplificatore operazionale a transconduttanza (OTA) con una caratteristica lineare ($G_m = \text{cost}$), la cui transconduttanza viene controllata variando la corrente di polarizzazione dell'OTA stesso. Tuttavia, non si riescono ad ottenere facilmente degli OTA con una bassa transconduttanza (a parità di costante di tempo, è necessaria un elevato valore di resistenza per poter tenere ridotto il valore di capacità) che godano di una buona linearità in un range di tensione esteso.

Il secondo problema dell'implementazione standard del monostabile, basata su porte logiche, è costituito dall'occupazione d'area. Infatti, valori di capacità e resistenza elevati sono difficilmente realizzabili in un circuito integrato. Oltre al fatto che, se il timer occupa molta area, allora il monitor diventa costoso e l'overhead di area rispetto al FIVR diviene meno giustificabile. Inoltre, quanto più grande l'area del monitor, tanto più aumenta la sua probabilità di guasto.

Una possibile soluzione per ovviare a questi problemi consiste nel sostituire il tradizionale monostabile con l'implementazione mostrata in figura 3.7, basata sull'idea esposta in [20]. Il circuito è composto di uno switch (che può essere implementato come un transistore nMOS) abilitato dal segnale di trigger di ingresso, un condensatore (che può essere implementato come un transistore nMOS con i terminali di source

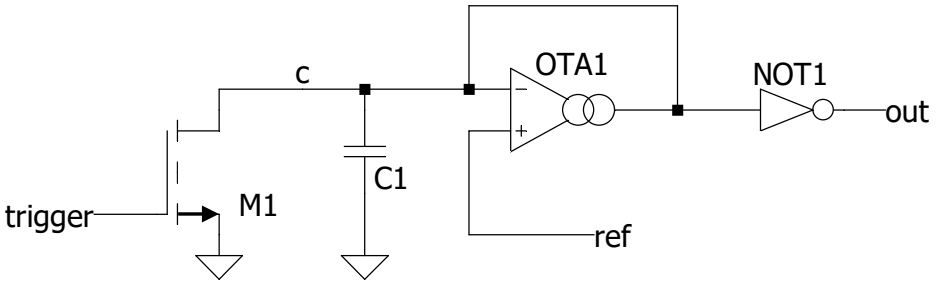


Figura 3.7: Implementazione alternativa del monostabile.

e drain cortocircuitati), un OTA ad alto guadagno e un inverter. La condizione di riposo del circuito prevede che il segnale di trigger sia a 0 (quindi, che lo switch sia aperto), la capacità sia carica alla tensione di riferimento V_{ref}[1] (in questo caso dato da $V_{in}/2$) e la tensione di uscita fornita dall'inverter sia nulla.

Quando all'ingresso di trigger si presenta un impulso, lo switch si chiude e scarica la capacità (idealmente in un tempo nullo). La tensione V_c passa istantaneamente da V_{ref} a 0, quindi l'uscita `out` si porta ad un valore logico alto. Dopo la commutazione della tensione V_c, la capacità comincia ad essere caricata dalla corrente di uscita dell'OTA. Poiché l'amplificatore è ad alto guadagno e la tensione differenziale di ingresso è positiva, la sua corrente di uscita è saturata al valore della corrente di bias ($I_o = I_B = \text{cost}$). In questo modo, la tensione ai capi del condensatore V_c aumenta linearmente verso la tensione di riferimento V_{ref}, poiché l'OTA è connesso ad inseguitore. Quando la tensione V_c supera la soglia logica V_{LT} dell'inverter, l'uscita commuta nuovamente a 0. Una volta che il condensatore è stato caricato, il circuito è ritornato alla condizione di riposo. L'inverter sarà dimensionato con la soglia logica più alta possibile, in modo da sfruttare la maggior parte possibile della durata della rampa generata dall'OTA. Nella realizzazione del circuito conviene aggiungere due inverters a valle di quello già presente in uscita in modo da rigenerare la forma d'onda, di cui il primo a soglia logica bassa ed il secondo a soglia logica alta.

Data la relazione costitutiva del condensatore

$$I_c(t) = C\frac{\mathrm{d}V_c(t)}{\mathrm{d}t} \tag{3.1}$$

[1] Da non confondere con il riferimento dato dalla PCU.

si ottiene facilmente

$$V_c(t) = V_c(0) + \frac{1}{C}\int_0^t I_c(\xi)\,\mathrm{d}\xi \tag{3.2}$$

da cui si evince che la tensione ai capi del condensatore cresce linearmente (dato che la capacità integra una corrente costante) e da cui segue che la durata della rampa data da V_c è pari a

$$T_c = \frac{V_{ref}C}{I_B} \tag{3.3}$$

Detta T_{out} la durata dell'impulso generato dall'inverter di uscita, si può scrivere la proporzione

$$\frac{V_{LT}}{T_{out}} = \frac{V_{ref}}{T_c} \tag{3.4}$$

da cui segue

$$T_{out} = \frac{V_{LT}}{V_{ref}} \cdot T_c = \frac{V_{LT}C}{I_B} \tag{3.5}$$

Dunque, la durata dell'impulso generato dal monostabile dipende dalla soglia logica dell'inverter (che però non è controllabile con precisione), dal valore di capacità (che non può essere troppo grande, come discusso sopra) e dalla corrente di bias dell'OTA, mentre è indipendente dalla tensione di riferimento. In particolare, la proporzionalità inversa rispetto alla corrente di bias sembrerebbe un punto a favore, perché maggiore la durata dell'impulso desiderata, minore la corrente assorbita dall'alimentazione. Tuttavia, si deve considerare che al diminuire della corrente di bias, si riduce anche il guadagno dell'OTA, per cui vi è il rischio che la corrente di uscita non sia più saturata per tensioni differenziali di ingresso sempre più prossime a zero. Questo farebbe sì che la corrente di uscita si riduca progressivamente deformando la rampa di V_c, la cui pendenza diminuirebbe.

In figura 3.8 è riportata la simulazione del monostabile effettuata con modelli ideali allo scopo di meglio dimostrarne il funzionamento. Si vede che in corrispondenza dell'impulso di trigger a circa $10ns$, la tensione V_c commuta immediatamente da V_{ref} a 0 e di conseguenza l'uscita dell'inverter si porta al valore logico alto. L'OTA comincia a caricare la capacità, la cui tensione sale linearmente. Superata la soglia logica V_{LT} dell'inverter, questo commuta nuovamente a 0, in corrispondenza dei $512ns$ di simulazione. Una volta caricato il condensatore, il circuito rimane nella condizione di riposo fino a quando non viene riapplicato un nuovo impulso di trigger in ingresso.

Nel caso in cui l'OTA presentasse un offset V_{OS}, la tensione V_c avreb-

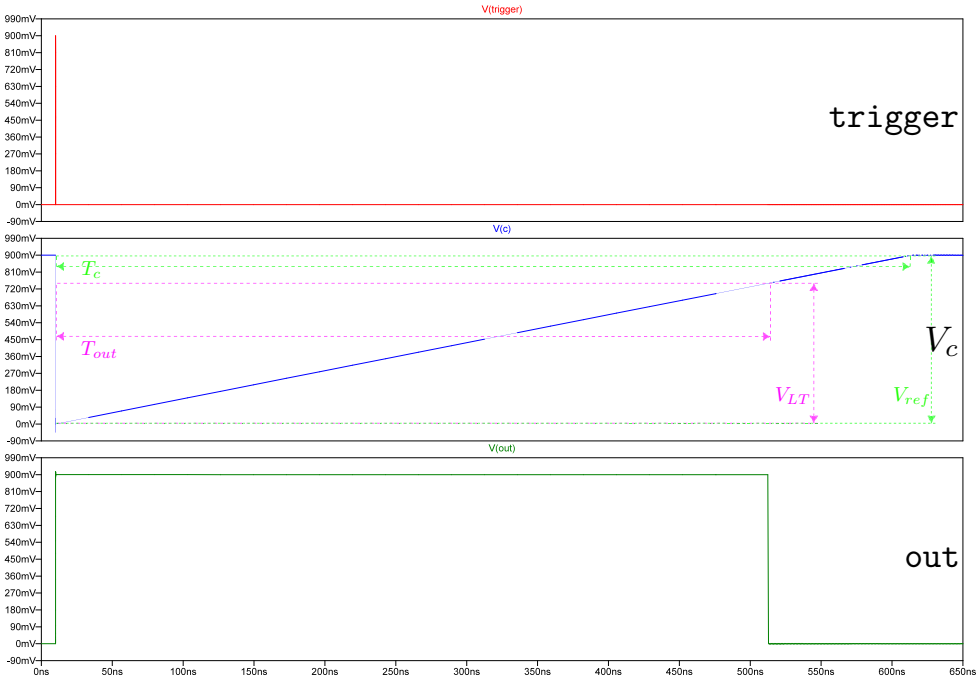


Figura 3.8: Simulazione dell'implementazione ideale del monostabile.

be un valore diverso nella condizione di riposo. Infatti, in tale situazione la tensione differenziale ai capi dell'OTA è nulla, quindi

$$V_d = V_{in}^+ - V_{in}^- + V_{OS} = 0 \tag{3.6}$$

da cui $V_c = V_{ref} + V_{OS}$. In realtà, questa non idealità dell'OTA non richiede nemmeno di essere compensata variando opportunamente la corrente di bias, perché la pendenza della rampa non viene modificata. Per avere un errore sulla durata T_{out} dell'impulso, la tensione di offset dovrebbe essere negativa e con un'ampiezza tale che la tensione V_c a riposo sia prossima alla soglia logica V_{LT} dell'inverter, ma questa situazione è decisamente inverosimile.

Incidentalmente, questa versione del monostabile è anche retriggerabile, cioè se dopo un iniziale impulso di trigger se ne presentasse un secondo, il circuito inizierebbe nuovamente il transitorio per generare l'impulso di durata prescritta, indipendentemente da quanto è durato l'impulso precedente.

Per la realizzazione del timer interno al monitor, analogamente a quanto fatto con la tradizionale implementazione, sarà necessario duplicare il monostabile (aggiungendo un ulteriore inverter alla replica) in modo tale da generare separatamente i segnali RS e $\overline{\texttt{RS}}$. Assumendo la

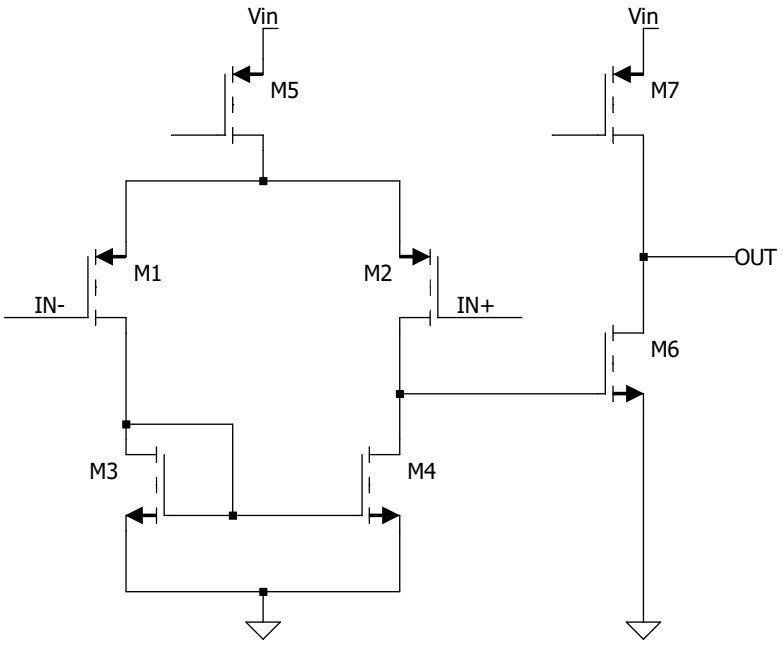


Figura 3.9: OTA di Miller (senza ramo di polarizzazione).

convenzionale ipotesi di guasti singoli, se la durata dell'impulso generata da un monostabile dovesse essere diversa da quella dell'impulso contrario generato dalla replica, allora l'errore verrebbe segnalato dall'EI.

Possibile implementazione dell'OTA

Di seguito si accenna ad una possibile implementazione dell'OTA sulla base di alcune considerazioni inerenti al circuito.

Dato il ridotto range di alimentazione (al massimo pari a $V_{in} = 1.8V$), si potrebbe scegliere di adottare l'OTA di Miller, mostrato in figura 3.9, costituito da una coppia differenziale cui segue uno stadio a source comune. Con questa topologia è possibile ottenere guadagni elevati senza compromettere eccessivamente lo swing dei segnali di ingresso e di uscita (infatti non è presente un transistore a gate comune in configurazione a cascode come nel caso dell'amplificatore Folded Cascode). Poiché il riferimento al terminale positivo è pari a metà del range di alimentazione e la tensione al morsetto negativo (che coincide con l'uscita) assume valori tra l'alimentazione negativa (ground) e il riferimento, si è utilizzata la coppia differenziale con gli ingressi sui transistori pMOS.

Il guadagno differenziale statico è dato da

$$A_{v_0} = g_{m_{in}} R_{out_{diff}} g_{m_{SC}} R_{out_{SC}} = \frac{g_{m_{in}}}{g_{d_2} + g_{d_4}} \cdot \frac{g_{m_{SC}}}{g_{d_6} + g_{d_7}} \qquad (3.7)$$

Poiché $g_m \propto \sqrt{I_D}$ e $g_d \propto I_D$, complessivamente si ottiene che $A_{v_0} \propto 1/I$, come anticipato alla sezione precedente. Quindi, dato che per ottenere un impulso di lunga durata è necessario ridurre la corrente erogata dall'OTA al condensatore, si può diminuire la corrente erogata dal generatore $M7$ aumentando indipendentemente la corrente erogata da $M5$ in modo da raggiungere il guadagno desiderato. Le tensioni di gate dei due generatori di corrente saranno ottenute da un opportuno ramo di polarizzazione. I transistori condivideranno il medesimo dimensionamento, ma, per ottenere diversi valori di corrente, avranno un diverso numero di *fingers*, in modo da migliorare il matching.

I poli dell'OTA di Miller sono ai nodi di uscita dei due stati, quindi le corrispettive pulsazioni sono date da:

$$\omega_{p_{diff}} = \frac{1}{R_{out_{diff}} C_{L_{diff}}} \quad \omega_{p_{SC}} = \frac{1}{R_{out_{SC}} C_{L_{SC}}} \tag{3.8}$$

Tipicamente il polo dominante è $\omega_{p_{diff}}$, perché la capacità di carico dello stadio differenziale è più elevata di quella dello stadio a source comune. Infatti, a causa dell'effetto Miller, la capacità $C_{gd_{SC}}$ viene moltiplicata circa per il guadagno dello stadio a source comune, per cui $C_{L_{diff}} = C_{gd_{SC}}(1 + A_{v_{SC}})$. Tuttavia, in questo caso il polo dominante è dato dal nodo di uscita dell'amplificatore a source comune dato che la sua capacità di carico è quella connessa al morsetto invertente dell'OTA. Tale capacità, per quanto abbia un valore contenuto per le motivazioni di cui sopra, avrà un valore sicuramente più elevato delle capacità parassite su quel nodo, in modo tale che sia dominante rispetto ad esse. Oltre al fatto che il guadagno statico dell'amplificatore a source comune in questo caso non è troppo elevato, dato che la corrente di bias deve essere sufficientemente piccola per garantire la durata dell'impulso desiderata. Quindi, in questo caso non è nemmeno necessario aggiungere la capacità di compensazione tra i due stadi, perché la compensazione viene effettuata direttamente dal carico, semplificando notevolmente la progettazione.

Capitolo 4

Efficacia della soluzione proposta

In questo capitolo viene dimostrato il funzionamento del monitor proposto attraverso i risultati di alcune simulazioni esemplificative, in modo da valutarne l'efficacia nel rivelare i guasti che possono affliggere il FIVR durante il suo funzionamento sul campo. Tutte le simulazioni di seguito riportate sono state effettuate in condizioni nominali, eccetto dove espressamente indicato.

4.1 Rivelazione di tensione di uscita errata

In figura 4.1 viene riportata una simulazione effettuata che parte da una situazione fault-free, in cui improvvisamente si assiste alla comparsa di un guasto che porta la tensione di uscita del FIVR fuori dal margine di tolleranza di 1 LSB. Il periodo di clock del segnale che viene dato in ingresso alle due VCDLs è pari a $100ns$ per una migliore lettura dell'immagine, data la durata complessiva della simulazione di $5.4\mu s$. All'inizio della simulazione, la tensione di uscita del FIVR coincide con il riferimento prescritto dalla PCU ed è pari al minimo del range erogabile, cioè $V_{out} = V_{ref} = 0.6V$. Durante questa condizione, che permane fino ai $200ns$, si può osservare come le uscite dell'EI `(ERR1,ERR2)` siano sempre uguali e si alternino alla frequenza del clock, segnalando il corretto funzionamento del circuito. A $200ns$ sul segnale `sample_ref` si osserva un impulso di breve durata, ad indicare il cambio di riferimento da parte della PCU. Effettivamente, in corrispondenza di tale impulso, la tensione di riferimento rigenerata dal DAC commuta alla tensione massima del range, cioè $V_{ref} = 1.1V$. Contemporaneamente il timer genera un impulso di reset di durata pari $502ns$ (nella figura è mostrato solo il segnale in forma vera). Come spiegato nel capitolo precedente, tale se-

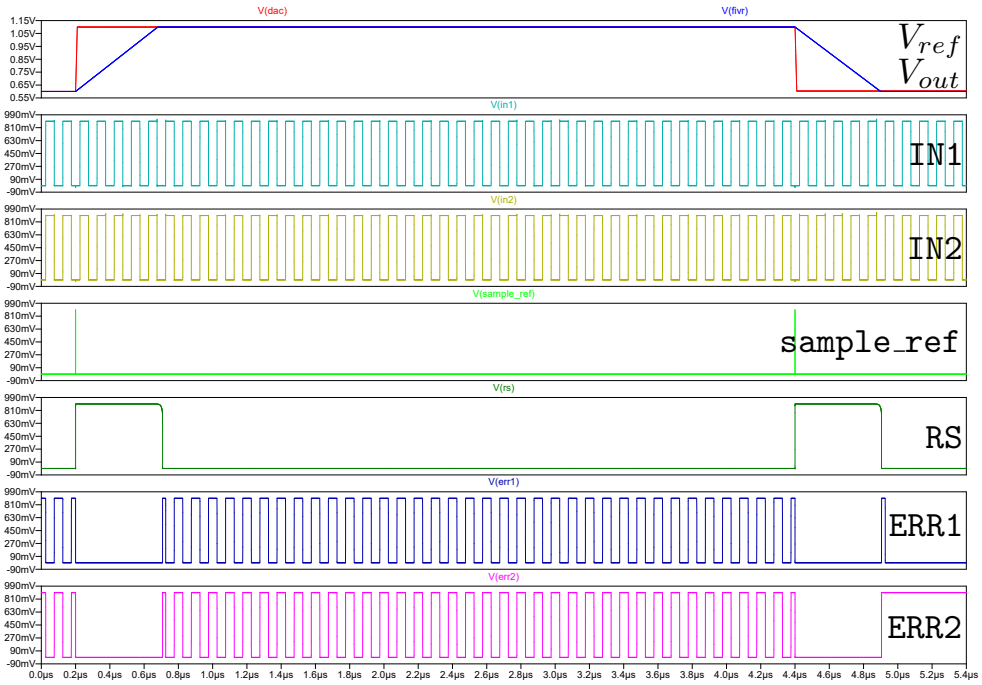


Figura 4.1: Simulazione del monitor per verificare la rivelazione di un guasto sul FIVR.

gnale di reset ha lo scopo di inibire l'EI durante il cambio di riferimento. Infatti, a seguito del cambio di riferimento da parte della PCU, l'uscita del FIVR va incontro ad un transitorio di una certa durata prima che la tensione si assesti al valore corretto. In questo lasso di tempo, l'EI deve essere obbligatoriamente inibito onde evitare che dia false indicazioni d'errore. Mentre `RS=1`, entrambe le uscite dell'EI sono forzate a `(ERR1,ERR2)=(0,0)`. Poiché i transistori operano in condizioni nominali, quindi sono esenti da fenomeni di aging, il transitorio del FIVR termina prima della fase di reset. Quando termina il reset (`RS=0`), le uscite dell'EI riprendono ad alternarsi alla frequenza del segnale di clock dato in ingresso alle due VCDLs con valori logici concordi, ad indicare il corretto funzionamento del circuito. Viene così dimostrato anche il soddisfacimento del requisito per cui deve esistere almeno una configurazione d'ingresso, corrispondente ad una parola di codice, per ciascuna delle due indicazioni di uscita corretta, cioè $(0, 0)$ e $(1, 1)$. In corrispondenza dei $4.4\mu s$ la PCU cambia nuovamente il riferimento, come si può notare dall'impulso sul segnale `sample_ref`, e la tensione rigenerata dal DAC commuta a $V_{ref} = 0.6035V$. Il comportamento del circuito è analogo a quanto visto alla commutazione precedente: il segnale di `sample_ref` abilita il monostabile e l'EI viene temporaneamente inibito. Tuttavia,

al termine del transitorio, l'uscita del FIVR si assesta a $V_{out} = 0.6000V$. La differenza fra le due tensioni è uguale all'ampiezza di 1 LSB, cioè $|V_{ref} - V_{out}| = 3.5mV$, quindi l'EI fornisce correttamente un'indicazione d'errore, poiché la sua sensibilità è tale da apprezzare la differenza $|d_1 - d_2|$ tra i ritardi dei segnali `IN1` e `IN2` in uscita dalle VCDLs. Si può notare come l'EI memorizzi l'indicazione d'errore `(ERR1,ERR2)=(0,1)` indipendentemente dal fatto che i suoi ingressi continuino a commutare.

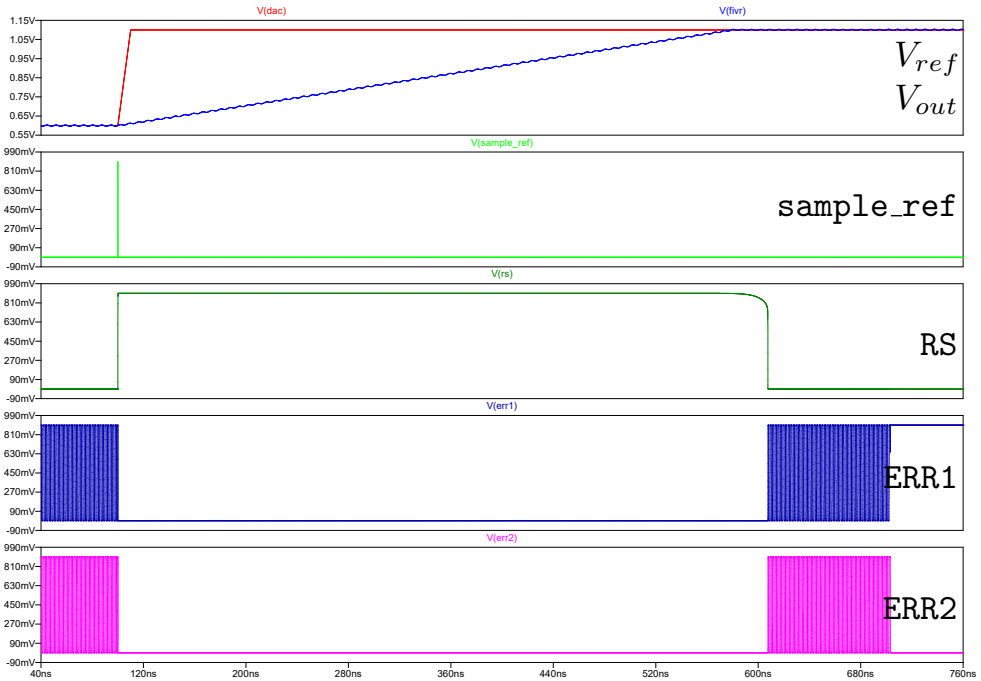
4.2 Rivelazione di ripple sulla tensione di uscita

In figura 4.2a viene mostrata una simulazione che dimostra l'abilità da parte del monitor di rivelare ripple eccessivo sulla tensione fornita dal FIVR. Il periodo di clock del segnale dato in ingresso alle VCDLs è pari a $2ns$, poiché per la rivelazione del ripple è necessario avere una frequenza di clock sufficientemente alta, a scapito della leggibilità della figura, di cui tuttavia viene proposto uno zoom in figura 4.2b.

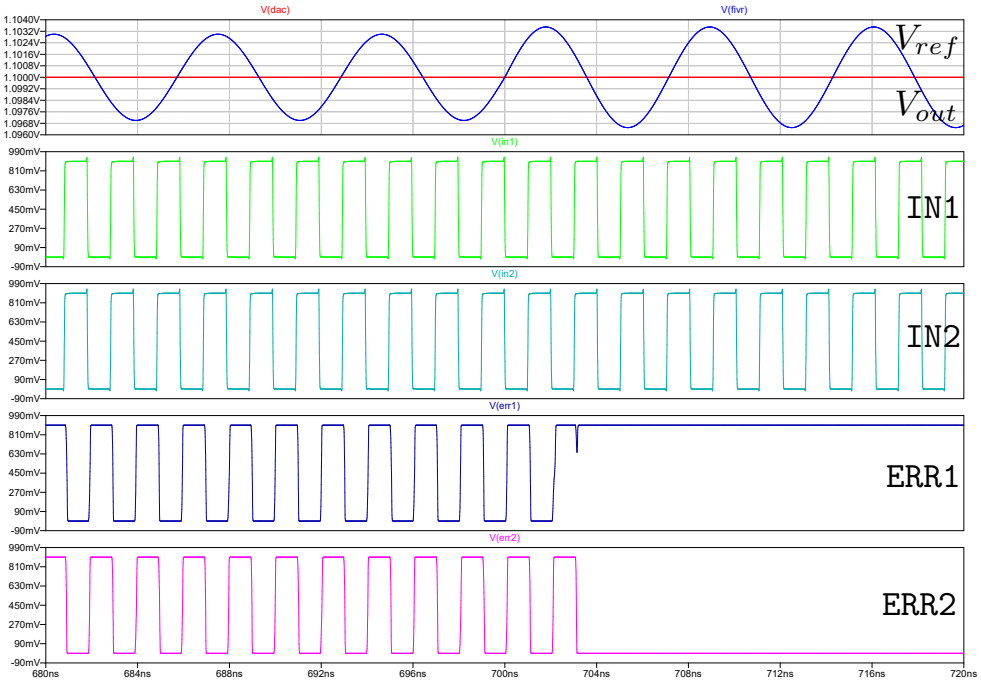
All'inizio della simulazione, la tensione di uscita del FIVR coincide con il riferimento rigenerato dal DAC nel valore medio ($\langle V_{out} \rangle = V_{ref} = 0.6V$), ma presenta un ripple di $3mV$ alla frequenza di $140MHz$, cioè la frequenza del segnale periodico triangolare usato nel modulatore PWM. Poiché l'ampiezza di tale ripple è tale che V_{out} non esca dal margine di tolleranza, l'EI fornisce un'indicazione di corretto funzionamento. Analogamente al caso precedente, ai $100ns$ il segnale `sample_ref` presenta un impulso di breve durata, ad indicare il cambio della tensione di riferimento da parte della PCU, che prescrive una tensione di $1.1V$. Tale segnale abilita il monostabile che genera il segnale di reset per inibire l'EI durante tutto il transitorio del FIVR in modo da evitare segnalazioni di errore errate. Al termine del transitorio, quando il segnale di reset si è abbassato, il valore medio della tensione di uscita del FIVR è ancora pari alla tensione fornita dal DAC, quindi l'EI fornisce un'indicazione di uscita corretta. Tuttavia, in corrispondenza dei $700ns$ un guasto sul FIVR fa sì che quest'ultimo presenti un ripple sull'uscita pari a $3.5mV$, cioè all'ampiezza di 1 LSB, come meglio mostrato in figura 4.2b. Poiché lo schema di rivelazione è stato progettato e dimensionato in modo da fornire un'indicazione d'errore per una differenza di tensioni pari o superiore a 1 LSB, l'EI segnala il guasto e memorizza l'indicazione d'errore.

4.3 Rivelazione dell'aging

In figura 4.3 viene mostrato il risultato di una simulazione effettuata per verificare l'abilità da parte del monitor di rivelare il rallentamento dei



(a) Simulazione del monitor per verificare la rivelazione di eccessivo ripple.



(b) Zoom della figura 4.2a.

Figura 4.2

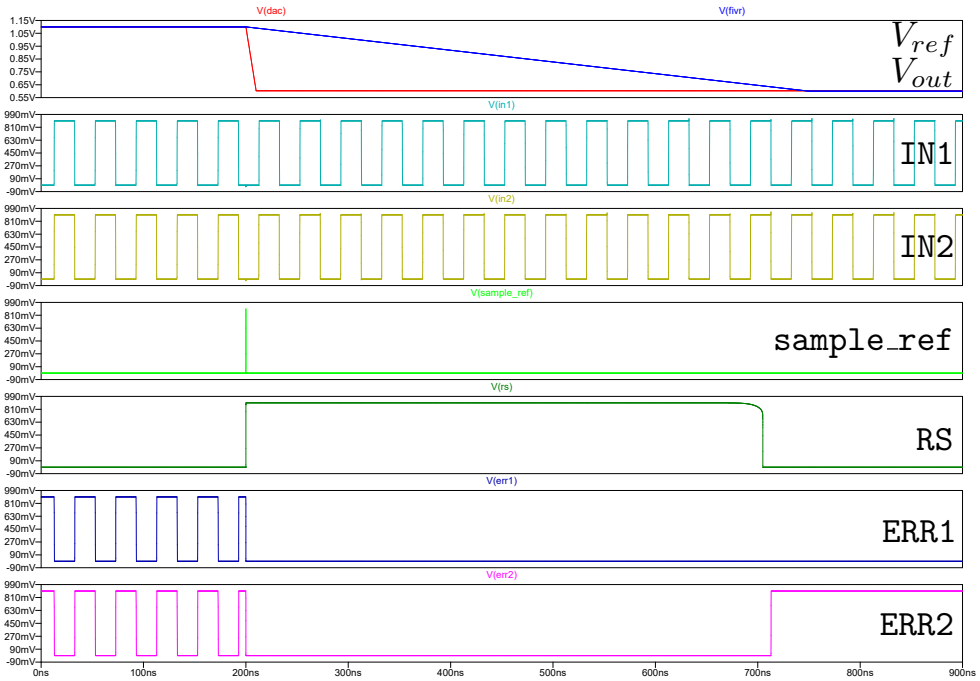


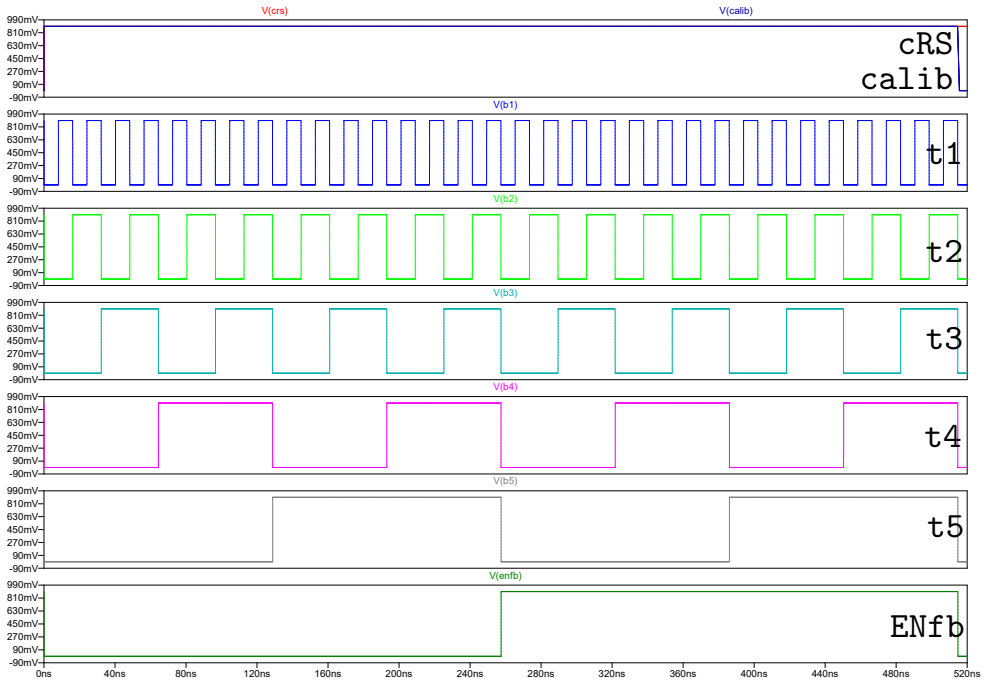
Figura 4.3: Simulazione del monitor per verificare la rivelazione del rallentamento dei transitori causato dall'aging.

transitori del FIVR causato da fenomeni di aging. Il periodo di clock del segnale dato in ingresso alle due VCDLs è pari a $20ns$ per una migliore lettura dell'immagine.

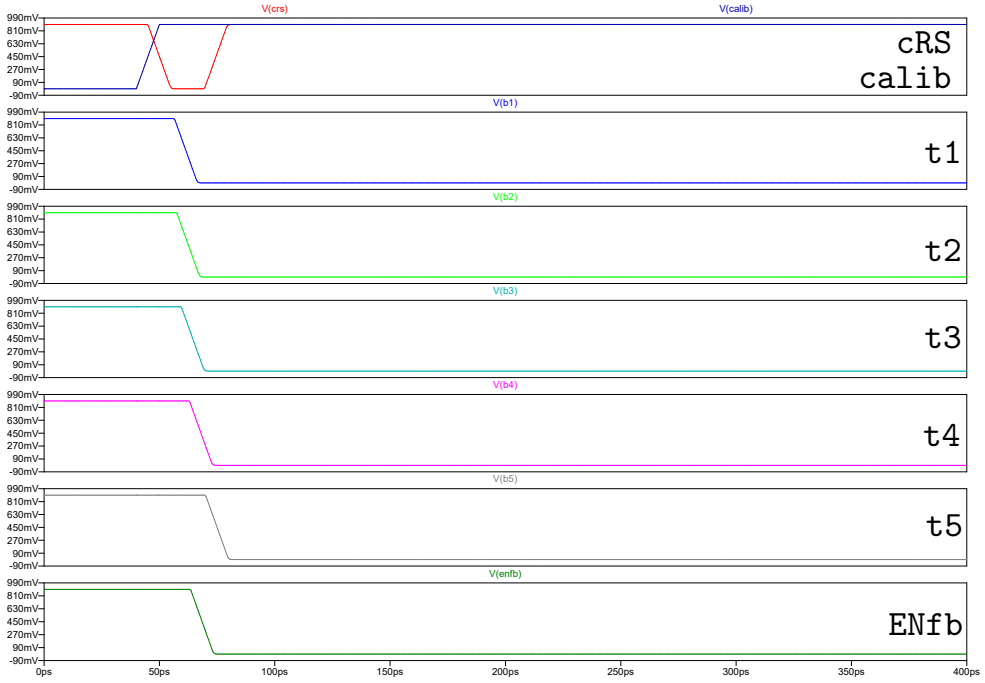
La simulazione mostra un cambio di riferimento da parte della PCU in corrispondenza dei $200ns$, come mostrato dall'impulso sul segnale `sample_ref`, che porta il FIVR a compiere la transizione di caso peggiore (di maggior durata), cioè da $V_{out} = 1.1V$ a $V_{out} = 0.6V$. In presenza di fenomeni di aging, si assiste ad un rallentamento dei transitori del FIVR. In questa simulazione è mostrato come, terminata la fase di reset allo scopo di inibire l'EI, la tensione di uscita del FIVR si debba ancora portare entro il margine di tolleranza di 1 LSB. Questo fa sì che, alla prima transizione utile dei segnali di ingresso `(IN1,IN2)` dell'EI (in questo caso una transizione di discesa), l'EI riveli l'errore e mantenga memorizzata la configurazione `(ERR1,ERR2)=(0,1)`.

4.4 Calibrazione dell'EI

Per dimostrare il corretto funzionamento del contatore, vengono riportate in figura 4.4a alcune forme d'onda di una delle simulazioni effettuate. La condizione iniziale della simulazione è meglio mostrata nello zoom



(a) Simulazione per verificare il funzionamento del contatore.



(b) Zoom della figura 4.4a.

Figura 4.4

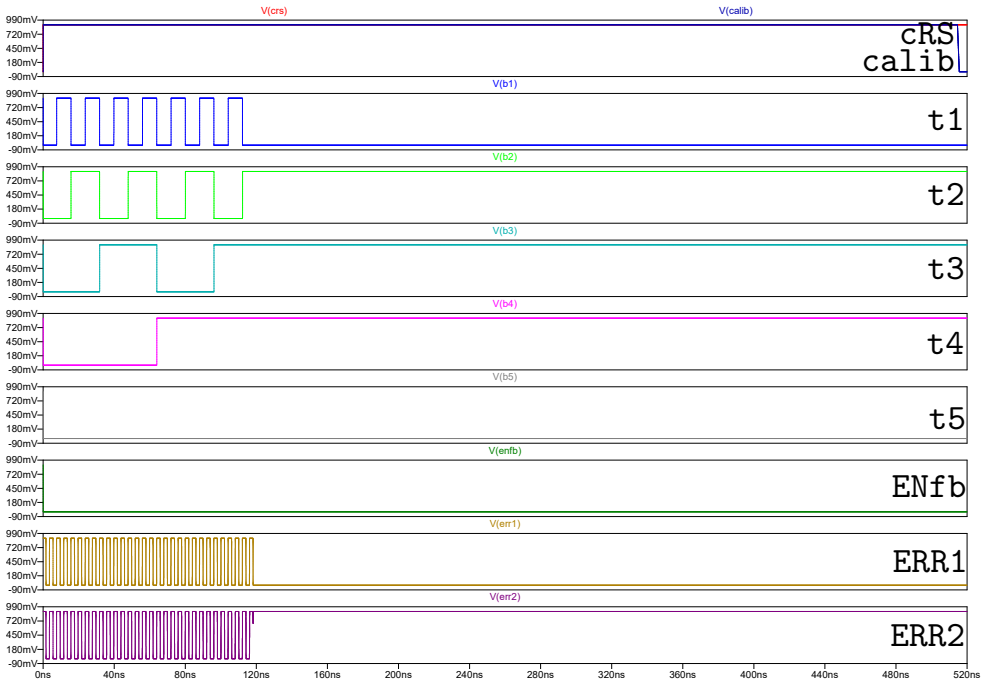


Figura 4.5: Simulazione per verificare il funzionamento del circuito di calibrazione.

mostrato in figura 4.4b. Si è supposto che all'accensione tutti i bits di tuning (che sono mostrati solo in forma vera) fossero pari a 1 per dimostrare il funzionamento del reset. All'accensione del sistema il segnale di calibrazione `calib` viene portato a 1 ed il segnale di reset `cRS` generato dal NAND gate ha un breve impulso negativo che consente di resettare tutti i FFs che costituiscono il contatore portando a spegnere tutti i bits di tuning. Il conteggio parte, dunque, da 0 e raggiunge la configurazione finale corrispondente a 63, dopo la quale il contatore riparte da 0. A quel punto il segnale di calibrazione viene disattivato. Nel caso specifico di questa simulazione, poiché il periodo del segnale di clock dato in ingresso alle due VCDLs è pari a $4ns$, si ha che il segnale di calibrazione rimane attivo per $4n \cdot 2^7 = 512ns$.

Il funzionamento del circuito di calibrazione e del relativo processo è mostrato in figura 4.5. Le condizioni iniziali sono esattamente quelle del caso precedente (meglio visibili in figura 4.4b). Quando il segnale `calib` commuta a 1, il LSB della parola di riferimento fornita dalla PCU viene complementato e fornito al DAC in modo che V_{out} e V_{ref} differiscano di 1 LSB. Partendo dalla configurazione corrispondente a 0, tutti i bits di tuning dell'EI sono spenti, quindi l'EI non è in grado di segnalare alcun errore, per cui `(ERR1,ERR2)` sono complementari e si alternano alla fre-

quenza del clock. L'AND gate fornisce il segnale di clock al primo FF dando inizio al conteggio. Una volta iniziato il conteggio, la conducibilità dei cammini di feedback dell'EI viene progressivamente incrementata. Tale processo dura fino a quando l'EI non riesce effettivamente a rivelare l'errore. Per la simulazione considerata, questo avviene poco prima dei $120ns$. L'indicazione d'errore fornita dall'EI inibisce il contatore, che memorizza la configurazione di bits di tuning che garantisce una sensibilità dell'EI pari a 1 LSB. Una volta che il segnale di calibrazione commuta a 0, il processo di calibrazione termina ed è possibile effettuare il reset dell'EI per passare al normale funzionamento del monitor. In questo modo, durante il normale funzionamento del monitor (quindi al di fuori della fase di calibrazione), nel caso fault-free in cui le uscite dell'EI sono uguali e si alternano alla frequenza del clock, `calib=0` garantisce che l'AND gate sia inibito, dunque che non vengano cambiati i bits di tuning.

Capitolo 5

Costi e self-checking ability

In questo capitolo verranno stimati i costi in termini di area del monitor proposto rispetto al FIVR e verrà discussa la self-checking ability dello schema di rivelazione stesso.

5.1 Valutazione dei costi e confronto con lo stato dell'arte

Un aspetto rilevante nella progettazione di circuiti integrati risulta essere l'occupazione d'area, a causa dei costi ad essa associati. Tuttavia, in ambito di alta affidabilità, vi è un motivo in più per minimizzare l'area occupata, che deriva dalla proporzionalità tra probabilità di guasto e l'area stessa. Di conseguenza, il monitor deve essere il più compatto possibile sia per minimizzare l'overhead di area rispetto al circuito da monitorare sia per minimizzare la probabilità di guasto rispetto al FIVR.

La stima dell'occupazione d'area viene effettuata in termini di *squares*, cioè il numero di transistori ad area minima necessari per l'implementazione dello schema di rivelazione. Nel caso di questo lavoro, in cui è stata utilizzata una tecnologia a $22nm$, l'area minima corrisponde chiaramente a $22nm \times 22nm$. Utilizzare gli squares come unità di misura consente di confrontare le aree indipendentemente dal nodo tecnologico, poiché non si considera la dimensione fisica del transistor, bensì il numero di transistori ad area minima.

Viene, infine, riportato il confronto tra l'area occupata dal monitor proposto in questa tesi rispetto all'occupazione d'area dello schema di rivelazione suggerito nello standard ISO 26262.

5.1.1 Occupazione d'area del monitor

Per fare un conteggio dei transistori presenti all'interno di ciascun blocco funzionale dello schema di rivelazione proposto, è utile calcolare innan-

Porta logica	Area occupata $[sq]$
NOT	3
NAND	8
NOR	10
AND	$8 + 3 = 11$
OR	$10 + 3 = 13$
XOR[1]	12
latch[2] C^2MOS with NAND reset	20

Tabella 5.1: Occupazione d'area delle singole porte logiche.

zitutto l'area occupata dalle singole porte logiche (tabella 5.1), che sono state dimensionate per essere simmetriche. Per la tecnologia considerata la mobilità degli elettroni è circa doppia rispetto a quella delle lacune ($\mu_n \approx 2\mu_p$), quindi la larghezza di canale dei transistori pMOS W_p sarà doppia rispetto a quella dei transistori nMOS W_n. Tutte le porte logiche sono implementate in topologia FCMOS.

Per quanto riguarda gli amplificatori operazionali (presenti solamente nel DAC), dato lo swing di alimentazione ridotto, si è ipotizzata un'implementazione secondo la topologia di Miller, anche denominata a due stadi, cui viene aggiunto uno stadio a drain comune per abbattere la resistenza di uscita in modo da poter pilotare carichi resistivi. La stima dell'area è stata effettuata in [5] a partire da [21].

Risulta impossibile una stima dell'area occupata da resistori e condensatori. Nei circuiti integrati, i resistori sono tipicamente implementati come transistori polarizzati in regione lineare, ma la resistenza così ottenuta non gode di una grande linearità, oppure tramite piste realizzate con materiali aventi un'elevata resistività. In questo caso, il valore di resistenza viene stimato come:

$$R = \frac{\rho}{T} \cdot \frac{L}{W} = R_{\square} \frac{L}{W} \tag{5.1}$$

dove ρ è la resistività del materiale, T è lo spessore della pista, L è la sua lunghezza e W la larghezza. A partire dalla resistività e dallo spessore, si definisce la *resistenza quadro* $R_{\square}$ (anche denominata R_{sq}), che è una grandezza dipendente dal processo. Poiché per la tecnologia utilizzata non è noto tale parametro, non è possibile effettuare una stima dell'area

[1] Da notare che l'implementazione a livello elettrico richiede i segnali di ingresso in forma vera e negata, quindi potrebbe essere necessario aggiungere fino a due NOT gates.

[2] Qui impropriamente annoverato fra le porte logiche.

Blocco	Area [sq]
DAC[3]	110
VCDLs	600
Timer[3]	63
EI	348
Calibrazione	302
Totale	1423

Tabella 5.2: Stima dell'occupazione d'area del monitor.

occupata dai resistori.

I condensatori vengono implementati nei circuiti integrati come transistori MOS con i terminali di source e drain cortocircuitati. La capacità ottenuta dipende dallo spessore e dal materiale usato per il dielettrico di gate dei transistori. La tecnologia utilizzata per le simulazioni usa il diossido di Silicio, quindi non è possibile effettuare una stima verosimile dell'area occupata dai condensatori. Infatti, le attuali tecnologie scalate hanno sostituito il diossido di Silicio con altri materiali, aventi costante dielettrica relativa maggiore in modo da ridurre ulteriormente le dimensioni dei dispositivi.

In tabella 5.2 viene riportata la stima dell'area occupata dal monitor, suddivisa per blocchi.

5.1.2 Confronto con soluzione dello standard

Lo schema di rivelazione proposto in questa tesi viene ora confrontato con la soluzione suggerita dallo standard ISO 26262, esaminata alla sezione 1.6. Seppure sul piano teorico lo schema suggerito dallo standard goda di un funzionamento molto semplice, dal punto di vista della implementazione pratica questo risulta in difetto rispetto al monitor progettato in questa tesi.

Innanzitutto, per quanto concerne le prestazioni, il rivelatore suggerito dallo standard, essendo basato su un ADC, risulta in grado di rivelare tutti i guasti critici, analogamente al monitor oggetto di questa tesi, tuttavia non è possibile rivelare il rallentamento dei transitori del FIVR dovuto ai fenomeni di aging. Inoltre, non è a priori garantita l'abilità self-checking nello schema di rivelazione suggerito dallo standard.

In seconda istanza, il monitor sviluppato in questa tesi risulta molto più vantaggioso in termini di area occupata, quindi di costi e di affidabilità del monitor stesso. In particolare, utilizzare un ADC (anziché un DAC) comporta un'occupazione d'area estremamente maggiore a causa

[3]Escluso il contributo di resistenze e condensatori.

Circuito	Area [sq]	Area overhead [%]
FIVR[4]	$\sim 4.5 \times 10^4$	-
Monitor proposto dallo standard	$\sim 3.5 \times 10^6$	7800
Monitor progettato in questa tesi	1423	3.16

Tabella 5.3: Confronto tra il monitor proposto e quello suggerito dallo standard.

dell'elevata frequenza di funzionamento (per poter rivelare il ripple) con una precisione di nove bits.

Il risultato del confronto è riassunto in tabella 5.3. Il monitor progettato in questa tesi si dimostra più efficace nella rivelazione dei guasti e assai meno costoso rispetto a quello suggerito dallo standard. L'overhead di area rispetto al circuito da monitorare, cioè il FIVR, risulta pressoché trascurabile.

Aggiungere il monitor di guasti rende il FIVR self-checking con un minimo overhead di area, conformandolo allo standard di safety ISO 26262. Il circuito risulta applicabile a tutte quelle applicazioni caratterizzate da esigenze di elevata reliability, cosiddette *safety critical*, come discusso nell'introduzione.

5.2 Analisi della self-checking ability

L'analisi dei guasti che possono affliggere il monitor e dei relativi effetti risulta particolarmente importante in ambito di alta affidabilità, poiché anche il monitor preposto alla rivelazione degli (eventuali) guasti che possono affliggere il blocco funzionale può guastarsi, seppure generalmente la probabilità che occorra un guasto nel monitor sia inferiore alla probabilità che il guasto si verifichi nel circuito da monitorare, data la minore complessità ed occupazione d'area. I guasti che possono (eventualmente) verificarsi all'interno del monitor potrebbero banalmente far sì che il monitor fornisca sempre un'indicazione di tensione di uscita del FIVR corretta, anche se questa sta al di fuori del margine di tolleranza. Si rende, dunque, necessaria un'analisi completa dei possibili guasti che possono affliggere il monitor, poiché quest'ultimo costituisce il blocco critico.

[4]Include l'area necessaria per realizzare i condensatori nei vari blocchi, ma non C_{out}.

Sono stati considerati gli stessi modelli di guasto utilizzati per l'analisi effettuata sul FIVR in [5]. Le ipotesi sono le stesse generalmente assunte per il funzionamento dei SCCs.

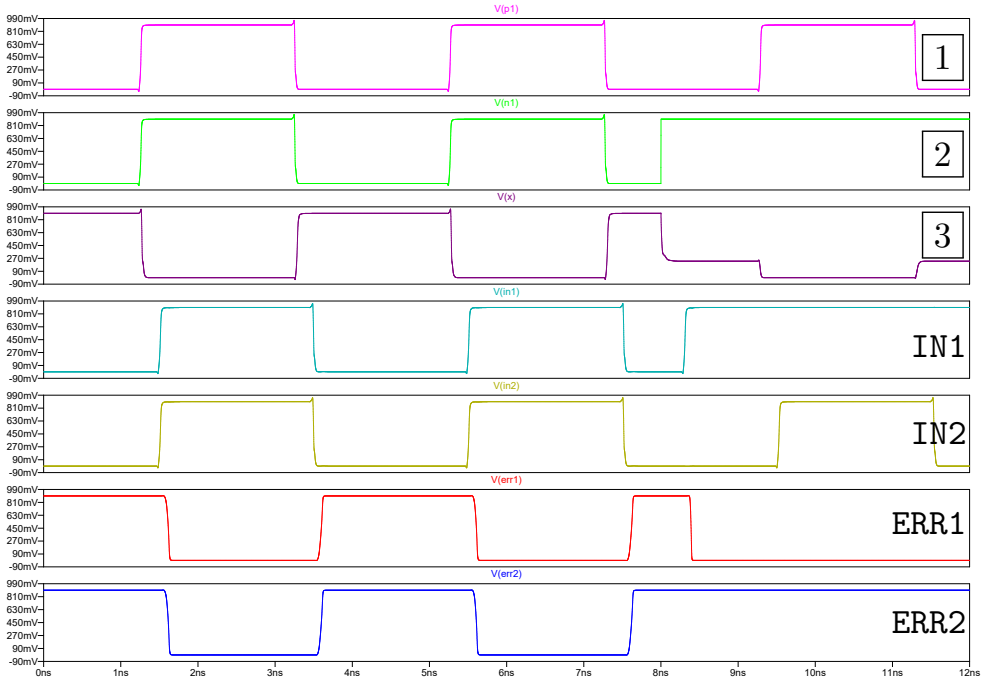
5.2.1 Transistor Stuck-on

Il modello di guasto di tipo SON è descritto in dettaglio alla sezione 1.2.2. Di seguito verrano brevemente valutati gli effetti che questo guasto può provocare nel funzionamento del monitor proposto. I transistori che possono essere soggetti a questo tipo di guasto si possono trovare in ogni blocco all'interno del rivelatore:

- nel DAC, sono i transistori che vengono pilotati dai bits per generare una corrente proporzionale alla significatività del bit stesso
- nelle VCDLs, sono i transistori che costituiscono l'inverter nella singola cella di ritardo
- nel timer, sono i transistori che costituiscono le porte NOR3 e NOT
- nell'EI, sono i transistori che costituiscono i vari gates (inverters, buffers tri-state, reti di feedback)
- nel circuito di calibrazione, sono i transistori che costituiscono le porte logiche e i FFs

Il guasto è stato simulato imponendo una tensione di gate tale da forzare l'accensione del transistor in esame.

Date le ipotesi circa il funzionamento dei SCCs, è facilmente intuibile come, vista la "simmetria differenziale" del circuito, il guasto venga sempre immediatamente rivelato. Si consideri, a titolo esemplificativo, che si verifichi un guasto di tipo SON al transistor nMOS di un inverter all'interno di una VCDL, come mostrato in figura 5.1. In tal caso, durante la fase alta del clock in ingresso all'inverter in esame, l'uscita si porta correttamente a zero, mentre, durante la fase bassa del clock, si accende anche il pMOS portando l'uscita ad un valore intermedio, che si traduce in un rallentamento per i transitori del gate logico a valle (cioè l'inverter della cella di ritardo seguente). In questo caso, indipendentemente dalla posizione della cella all'interno della linea di ritardo, l'EI rivela immediatamente il guasto, poiché i suoi ingressi avranno un diverso ritardo. Nell'esempio riportato, il guasto si verifica in corrispondenza degli $8ns$. La tensione di uscita dell'inverter si porta ad una tensione intermedia circa pari a $230mV$, cioè ben al di sotto della soglia logica. Questo fa sì che, indipendentemente che la fase del clock in ingresso



1 Tensione di gate del transistore pMOS appartenente all'inverter guasto.

2 Tensione di gate del transistore nMOS affetto dal guasto.

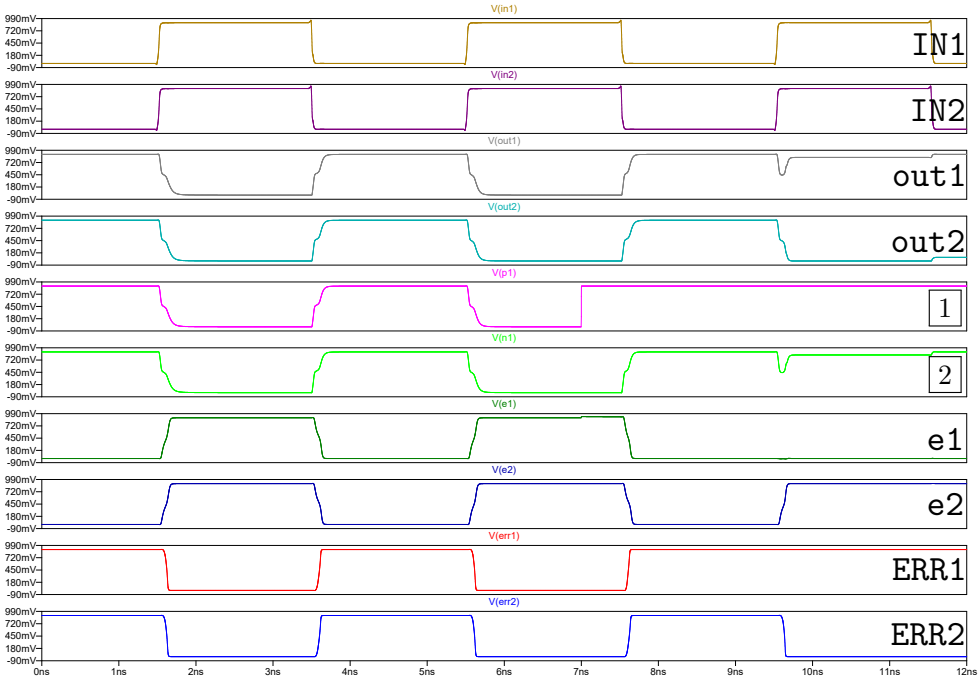
3 Tensione di uscita dell'inverter guasto.

Figura 5.1: Esempio di guasto SON ad un transistore nMOS di un inverter appartenente ad una VCDL.

all'inverter sia alta o bassa, il valore logico venga rigenerato già all'inverter successivo, che avrà sempre una tensione alta in uscita. Infatti, l'ingresso dell'EI `IN1` non commuta più dopo che si è verificato il guasto e l'EI rivela immediatamente il guasto, bloccando le proprie uscite a `(ERR1,ERR2)=(0,1)`.

5.2.2 Transistor Stuck-open

Il modello di guasto di tipo SOP è descritto in dettaglio alla sezione 1.2.3. Valgono le medesime considerazioni preliminari fatte alla sezione 5.2.2. Il guasto è stato simulato imponendo una tensione di gate tale da forzare lo spegnimento del transistor in esame. Ad esempio, si consideri un guasto SOP al transistore pMOS dell'inverter `NOT2` appartenente al ramo superiore dell'EI (quello che riceve l'ingresso `IN1`), come mostrato in figura 5.2. Nell'esempio, il guasto si verifica in corrispondenza dei $7ns$, cioè proprio mentre il transistore dovrebbe essere acceso, dunque l'uscita `e1` si porta in alta impedenza. Successivamente, quando `IN1`



1 Tensione di gate del transistore pMOS affetto dal guasto.

2 Tensione di gate del transistore nMOS appartenente all'inverter guasto.

Figura 5.2: Esempio di guasto SOP ad un transistore pMOS di un inverter `NOT2` interno all'EI.

commuta verso 0, si osserva che `out1` ha la transizione opposta e anche l'inverter guasto `NOT2` commuta correttamente portando l'uscita `e1` a 0. Questo corrisponde al vettore di inizializzazione necessario per portare l'uscita al valore logico opposto rispetto a quello dato dalla rete in cui è presente il transistore guasto (la rete in cui è presente il transistore guasto è quella di pull-up, che porterebbe l'uscita al valore logico alto in assenza di guasto), come era stato descritto alla sezione 1.2.3. L'EI non ha ancora rivelato il guasto, ma il monitor nel frattempo continua a funzionare correttamente, quindi viene preservata la data integrity. Il vettore di attivazione del guasto viene dato in corrispondenza della successiva transizione di salita del segnale `IN1`. L'uscita `out1` dell'inverter `NOT1` dovrebbe, infatti, portarsi a 0, tuttavia il cammino di feedback si attiva, poiché l'inverter `NOT2` non riesce ad iniziare la commutazione verso il valore logico alto, dato che il transistore pMOS è affetto da un guasto di tipo SOP. Quindi l'EI rivela il guasto, memorizzando `(ERR1,ERR2)=(1,0)`.

Per la rivelazione dei guasti di tipo SOP risulta particolarmente im-

portante il fatto che, nel caso fault-free, vi siano due indicazioni di correttezza dell'uscita che si alternano. In questo modo è sempre possibile avere un vettore di inizializzazione cui segue quello di attivazione per poter correttamente rivelare il guasto ed averne l'immediata segnalazione.

5.2.3 Bridging resistivo

Il modello di guasto di tipo BF è descritto in dettaglio alla sezione 1.2.1. Come riportato in tale sezione, sono stati considerati i valori di resistenza $R_B \in [0, 100k]\Omega$. In particolare, sono state effettuate simulazioni con i seguenti valori $[1k, 10k, 25k, 100k]\Omega$. Le considerazioni circa la rivelazione dei guasti sono analoghe a quelle fatte alla sezione 5.2.1, dato che il guasto di tipo SON può essere interpretato come un caso particolare del guasto di tipo BF.

Conclusione

Questo lavoro di tesi, svolto nell'ambito di un progetto di ricerca in collaborazione con *Intel Corporation*, è iniziato dall'analisi dello stato dell'arte riguardante il FIVR ed i fenomeni di aging e considerando i risultati conseguiti precedentemente in [4, 5].

L'attività principale è stata la progettazione di un nuovo schema di rivelazione dei guasti che non fosse affetto dalle problematiche di quello precedentemente ideato in [5]. Questo ha richiesto di cambiare completamente l'approccio al problema della rivelazione dei guasti, basando il principio di funzionamento del nuovo monitor non su un confronto diretto fra tensioni, bensì traducendo la differenza di tensione in una differenza di ritardo con una nuova strategia di on-line testing. Il monitor così progettato è in grado di rivelare i guasti critici del FIVR, tali da far sì che la sua uscita si porti al di fuori del margine di tolleranza, imposto uguale all'ampiezza di 1 LSB della parola di riferimento di nove bits fornita dalla PCU, cioè alla risoluzione del DAC presente nel FIVR. Il monitor è anche in grado di fornire un'indicazione d'errore qualora il ripple sulla tensione di uscita del FIVR abbia un'ampiezza eccessiva, tale da fuoriuscire dal margine di tolleranza, nonostante il valore medio possa essere corretto, nonché è in grado di rivelare il rallentamento dei transitori del FIVR dovuto a fenomeni di invecchiamento dei transistori, a seguito del cambio di riferimento da parte della PCU. Inoltre, lo schema di rivelazione risulta self-checking, cioè è in grado di autocollaudarsi rispetto ai possibili guasti interni. Questo risulta fondamentale in ambito di alta affidabilità per evitare di compromettere l'efficacia del monitor nel rivelare i guasti che possono verificarsi nel FIVR.

Concepito lo schema ad alto livello del monitor, si è passati all'implementazione a livello circuitale tramite LTSpice. Diverse simulazioni sono state effettuate per dimensionare correttamente il circuito e valutarne, infine, le prestazioni in termini di capacità di rivelare i guasti durante il normale funzionamento sul campo del FIVR. In particolare, per quanto riguarda il rallentamento dei transitori a causa di fenomeni di aging, si è considerato il transitorio di caso peggiore, cioè quello di una commu-

tazione in discesa pari alla massima escursione possibile per l'uscita del FIVR, cioè da 1.1 a $0.6V$. Trascorso il tempo necessario in condizioni nominali per effettuare questo transitorio, se la tensione di uscita del FIVR si trova al di fuori del margine di tolleranza, il monitor fornisce un'indicazione d'errore.

Dopo un'approfondita analisi preliminare tramite simulazioni Monte Carlo per capire l'incidenza e l'effetto delle PPVs sulle linee di ritardo (in termini di variazioni del ritardo reciproco fra i due segnali in uscita dalle VCDLs) e sull'EI (in termini di sensibilità) in maniera indipendente, si è modificato l'EI in modo da introdurre la possibilità di tuning. In questo modo, il monitor risulta robusto alle PPVs, poiché, qualsiasi siano le condizioni dello stesso, ad ogni accensione del sistema un opportuno circuito di calibrazione effettua in maniera automatica il tuning dell'EI in modo da garantire in ogni caso un margine di tolleranza pari a 1 LSB, come previsto in fase di progettazione. Questo consente di compensare anche gli effetti dei fenomeni di invecchiamento dei dispositivi (sia delle VCDLs sia dell'EI) e variazioni della temperatura d'esercizio, quindi delle condizioni operative.

Lo schema di rivelazione proposto in questa tesi costituisce una soluzione a basso costo per monitorare il funzionamento del FIVR. L'overhead di area risulta trascurabile, rendendolo una soluzione di gran lunga più economica rispetto a quanto suggerito dallo standard ISO 26262, rispetto al quale risulta anche più performante. L'aggiunta del monitor consente di rispettare i requisiti di safety come richiesto dallo standard, rendendolo idoneo all'utilizzo in applicazioni caratterizzate da esigenze di elevata reliability, come ad esempio l'autonomous drive.

Bibliografia

[1] Edward A Burton et al. "FIVR—Fully integrated voltage regulators on 4th generation Intel® Core™SoCs". In: *2014 IEEE Applied Power Electronics Conference and Exposition-APEC 2014*. IEEE. 2014, pp. 432–439.

[2] Bill Bowhill et al. "The Xeon® processor E5-2600 v3: A 22 nm 18-core product family". In: *IEEE Journal of Solid-State Circuits* 51.1 (2015), pp. 92–104.

[3] Noah Sturcken et al. "A switched-inductor integrated voltage regulator with nonlinear feedback and network-on-chip load in 45 nm SOI". In: *IEEE Journal of Solid-State Circuits* 47.8 (ago. 2012), pp. 1935–1945.

[4] Alex Menghi. "Affidabilità di Sistemi di Regolazione dell'Alimentazione per Processori Multi-Core". Laurea Magistrale. Alma Mater Studiorum — Università di Bologna, mar. 2019.

[5] Alessandro Stefani. "Progetto di Sistemi di Regolazione dell'Alimentazione ad Alta Affidabilità per Processori Multi-Core". Laurea Magistrale. Alma Mater Studiorum — Università di Bologna, lug. 2019.

[6] Bogdan Tudor et al. "MOS device aging analysis with HSPICE and CustomSim". In: *Synopsys, White Paper* (ago. 2011).

[7] Bogdan Tudor et al. "MOSRA: An efficient and versatile MOS aging modeling and reliability analysis solution for 45nm and below". In: *2010 10th IEEE International Conference on Solid-State and Integrated Circuit Technology*. IEEE. 2010, pp. 1645–1647.

[8] John Keane e Chris H Kim. "Transistor aging". In: *IEEE Spectrum* 48.5 (2011), pp. 28–33.

[9] James H Stathis e Sufi Zafar. "The negative bias temperature instability in MOS devices: A review". In: *Microelectronics Reliability* 46.2-4 (2006), pp. 270–286.

[10] Ying-Zu Lin et al. "A 9-bit 150-MS/s subrange ADC based on SAR architecture in 90-nm CMOS". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 60.3 (mar. 2013), pp. 570–581.

[11] *ISO 26262-11:2018, Road vehicles – Functional Safety – Part 11: Guidelines on application of ISO 26262 to semiconductors*. 2018.

[12] Arizona State University. *Predictive Technology Model*. 2008. URL: `http://ptm.asu.edu`.

[13] Jan M Rabaey, Anantha P Chandrakasan e Borivoje Nikolić. *Digital integrated circuits: a design perspective*. Vol. 7. Pearson education Upper Saddle River, NJ, 2003.

[14] Salah Hanfoug, Nour-Eddine Bouguechal e Samir Barra. "Behavioral non-ideal model of 8-bit current-mode successive approximation registers ADC by using Simulink". In: *International Journal of u-and e-Service, Science and Technology* 7.3 (2014), pp. 85–102.

[15] Bilal I Abdulrazzaq et al. "A review on high-resolution CMOS delay lines: towards sub-picosecond jitter performance". In: *SpringerPlus* 5.1 (2016), p. 434.

[16] Mohammad Maymandi-Nejad e Manoj Sachdev. "A digitally programmable delay element: design and analysis". In: *IEEE transactions on very large scale integration (VLSI) systems* 11.5 (2003), pp. 871–878.

[17] GS Jovanovic e MK Stojcev. "Vernier's delay line time–to–digital converter". In: *Scientific Publications of the State University of Novi Pazar, Ser. A: Appl. Math Inform. and Mech* 1 (2009), p. 1.

[18] Goran Jovanović e Mile K Stojčev. "Voltage controlled delay line for digital signal". In: *Facta universitatis-series: Electronics and Energetics* 16.2 (2003), pp. 215–232.

[19] Cecilia Metra, Luca Schiano e Michele Favalli. "Concurrent detection of power supply noise". In: *IEEE Transactions on Reliability* 52.4 (dic. 2003), pp. 469–475.

[20] Yu-Kang Lo e Hung-Chun Chien. "Current-controllable monostable multivibrator with retriggerable function". In: *Microelectronics Journal* 40.8 (2009), pp. 1184–1191.

[21] Andrea Stanco. "Progetto di un Amplificatore Operazionale a due stadi in tecnologia CMOS". Laurea. Università degli Studi di Padova, lug. 2011.